

Augustin's Method - Part II: The Sphere Packing Bound

Bariş Nakiboğlu

Canım halam Fatma Nakiboğlu Aydiç'in anısına adanmıştır.

Dedicated to the memory of my dear aunt Fatma Nakiboğlu Aydiç.

Abstract

The channel coding problem is reviewed for an abstract framework. If the scaled Renyi capacities of a sequence of channels converge to a finite continuous function φ on an interval of the form $(1 - \varepsilon, 1]$ for an $\varepsilon > 0$, then the capacity of the sequence of channels is $\varphi(1)$. If the convergence holds on an interval of the form $(1 - \varepsilon, 1 + \varepsilon)$ then the strong converse holds. Both hypotheses hold for large classes of product channels and for certain memoryless Poisson channels. A sphere packing bound with a polynomial prefactor is established for the decay rate of the error probability with the block length on any sequence of product channels $\{W_{[1,n]}\}_{n \in \mathbb{Z}^+}$ satisfying $\max_{t \leq n} C_{1/2, W_t} = O(\ln n)$. For discrete stationary product channels with feedback sphere packing exponent is proved to bound the exponential decay rate of the error probability with block length from above. The latter result continues to hold for product channels with feedback satisfying a milder stationarity hypothesis. A sphere packing bound with a polynomial prefactor is established for certain memoryless Poisson channels.

CONTENTS

I	Introduction	2
I-A	Main Contributions	3
I-B	The Channel Coding Problem	3
I-C	Product Channels, Stationarity and Memorylessness	4
I-D	Product Channels With Feedback	5
II	Operational Significance of The Renyi Capacity and Asymptotic Quantities	7
II-A	The Sphere Packing Exponent	7
II-B	Gallager's Inner Bound	8
II-C	Arimoto's Outer Bound	9
II-D	The Channel Capacity	10
III	The Sphere Packing Bound for Product Channels	12
III-A	Tilting with A Family of Measures	13
III-B	A Lower Bound on the Probability of Small Deviations via Berry Essen Theorem	15
III-C	Averaged Renyi Center, Averaged Renyi Capacity and Averaged Sphere Packing Exponent	16
III-D	Outer Bounds for Codes on Product Channels	19
III-E	Proof of Theorem 2	22
IV	The Sphere Packing Bound for Product Channels With Feedback	23
IV-A	A Taylor's Expansion For The Renyi Divergence	23
IV-B	Renyi Divergence Tradeoff for Tilting and The Dummy Channel Method	24
IV-C	An Outer Bound for Codes on Product Channels With Feedback	27
IV-D	Proof of Theorem 3	29
V	Discussion and Future Directions	29
	Appendix	29
A	Minimum Sigma-algebra for Information Transmission	29
B	The Sphere Packing Bound for Codes on Certain Poisson Channels	31
	References	33

I. INTRODUCTION

Most proofs establishing outer bounds for the channel coding problem under fixed rate, fixed error probability, or slowly vanishing error probability hypothesis rely on either a type based expurgation [3], [28], [47] or a distinction of cases based on types [40], [50]. Although similar bounds can, usually, be obtained using information spectrum approach [31] with greater generality, one has to give up the initial non-asymptotic bound in order to do so. This relative advantage of the method of types [14], [15] over information spectrum approach [27] emerges from four distinct assumptions: the product structure of the sample space, the product structure of the probability measures, finiteness of the input set, and the stationarity of the channel. Finite input set assumption and product structure assumptions can be removed and the stationarity assumption can be relaxed if one gives up the concept of type for the concept of typicality. Typicality arguments are, usually, employed in deriving asymptotic results, but they can also be used to obtain non-asymptotic bounds.

Augustin's proof of the sphere packing bound [9] stands out in this high level classification of the techniques for deriving outer bounds. It establishes a non-asymptotic bound without assuming finiteness of the input set or the stationarity of the channel. In this second paper of the two part series, our main aim is to build an understanding of Augustin's method around the concepts of the capacity and the center. We believe such an understanding can guide us when we apply Augustin's method to the other information transmission problems. To build such an understanding, we derive sphere packing bounds using Augustin's method in a way that makes the role of the Renyi capacities and the Renyi centers more explicit. In order to describe the operational significance of Renyi capacities more completely, we determine the channel capacity for a diverse class of channels using Gallager's inner bound [24] and Arimoto's outer bound [6].

Shannon, Gallager and Berlekamp [47, Theorem 2] published the first rigorous proof of the sphere packing bound for arbitrary discrete stationary product channels¹ (DSPCs). The following year Haroutunian [28, Theorem 2] published an alternative proof that holds for arbitrary stationary product channels (SPCs) with finite input alphabets. In [28], Haroutunian expressed the sphere packing exponent in an alternative form, which he proved to be equal to the one in [47]. The next year, Augustin [8, Theorem 4.7] published another proof of the sphere packing bound that holds for non-stationary product channels with possibly infinite input set. Augustin's sphere packing bound [8, Theorem 4.7a] holds even for product channels with infinite channel capacity. In the same paper, Augustin also established a sphere packing bound with a polynomial prefactor [8, Theorem 4.8], for a hypothesis that is satisfied by all DSPCs.

The first two proofs of the sphere packing were relying on expurgations based on the empirical type of the input codewords. As a result, proofs in [47] and [29] were only applicable to SPCs with finite input alphabets. Nevertheless, they had greater impact on the field than Augustin's proof. Variants of Haroutunian's proof can be found in [14], [15], [17], [30], [36]. For DSPCs, Haroutunian's method leads to a sphere packing bound with polynomial prefactor, i.e. a prefactor of the form $e^{-O(\ln n)}$. The prefactor of Shannon, Gallager, Berlekamp proof is $e^{-O(\sqrt{n})}$, which is considerably worse. It has been improved for the moderate block lengths in [51] and [54]. However, it is not clear from the analysis presented in [51] and [54], whether they lead to any improvements in the asymptotic decay rate of the prefactor or not.

Using the list decoding variant of Gallager's inner bound, one can see that the exponential decay rate of the sphere packing bound, i.e. sphere packing exponent, is tight. But determining the right prefactor for the sphere packing bound is still an open problem even for DSPCs. Altug and Wagner [1] considered DSPCs with positive transition probabilities satisfying certain symmetry conditions [25, p.94] and proved that sphere packing outer bounds hold for prefactor $n^{-\frac{1+\epsilon}{2\alpha}}$ for any $\epsilon > 0$ for certain α in $(0, 1)$. Their result is tight because they have proved in a later paper [4] that Gallager's inner bound can be improved to have a prefactor $n^{-\frac{1}{2\alpha}}$, for the α mentioned above, for arbitrary DSPCs. For general DSPCs, we only have bounds for the constant composition codes that are also due to Altug and Wagner [5].

Sphere packing bound has been conjectured to hold for codes on DSPCs with feedback. Assuming certain symmetries, Dobrushin [19] proved it to be the case. However, it was challenging to prove the conjecture for arbitrary DSPCs with feedback because of the reliance of the standard proofs on type based expurgations. In [29], Haroutunian established an outer bound for codes on arbitrary DSPCs with feedback, but Haroutunian's exponent is equal to the sphere packing exponent only for DSPCs with certain symmetries.² Augustin has shown that the sphere packing bound holds for arbitrary DSPCs with feedback, [9, Theorem 41.7]. Augustin conciliates the reader with a proof sketch rather than a complete proof. Nevertheless, his argument is sound and details can be added to the sketch, if needed.³ Few years after Augustin's manuscript [9], Sheverdyaev suggested another proof in [48]. We believe Sheverdyaev's proof is correct in essence, however it is supported rather weakly at certain critical points.⁴ Palaiyanur's thesis [39, A7] includes an in depth discussion of the subtleties of [48].

¹Recently, Dalai gave an account of earlier results and proofs in [16, Appendix B].

²There are other partial results [13], [37], [38] establishing the sphere packing bound for certain families of codes on DSPCs with feedback by curtailing the encoding schemes.

³Author has done so for DSPCs with non-zero transition probabilities in the fall of 2012 at UC Berkeley while working under the supervision of Anant Sahai. Later, author has confirmed that the details of Augustin's proof sketch can be filled for arbitrary DSPCs, as well.

⁴Sheverdyaev has two major results about DSPCs with feedback in [48]. Our reservations are about the result for the rates less than the order one Renyi capacity. Sheverdyaev proves his result for the rates greater than the order one Renyi capacity satisfactorily.

A. Main Contributions

- (I) Theorem 1 determines the channel capacity for a large class of sequences of channels and presents a sufficient condition for the existence of a strong converse. Theorem 1 seems to be new, but it merely answers the question “What does Gallager’s inner bound and Arimoto’s outer bound say about the channel capacity and the existence of a strong converse?” Characterizing channel capacity and the conditions for the existence of a strong converse, in general, is a separate issue that has already been addressed by Verdú and Han [52]. Poisson channels⁵ $\Lambda^{[T,a,b,\varrho]}$, $\Lambda^{[T,a,b,\leq\varrho]}$, $\Lambda^{[T,a,b,\geq\varrho]}$, and $\Lambda^{[T,a,b]}$ satisfy the sufficient condition for the strong converse given in Theorem 1. The strong converses for Poisson channels have been reported before in [12] and [53], but only for the zero dark current cases (i.e. $a = 0$ cases).
- (II) Theorem 2 proves the sphere packing bound with a polynomial prefactor for sequences of product channels without assuming the stationarity of the channels or the finiteness of the input sets. Theorem 2 does not even need the order one Renyi capacity of the channels in the sequence to be finite. Theorem 2 applies to all stationary product channels, hence it applies to the Poisson channels with bounded intensity functions described in [35, Example 10]. Unlike the sphere packing bounds derived by Augustin in [8] and [9], Theorem 2 does not rely on the uniform continuity hypothesis described in Assumption 2. Furthermore, Theorem 2 establishes the sphere packing bound with a polynomial prefactor for an hypothesis considerably weaker than the one assumed by Augustin for a similar result [8, Theorem 4.8]. A more detailed comparison of Theorem 2 and Augustin’s results are presented on page 12.
- (III) Theorem 3 establishes the sphere packing exponent as an upper bound to the decay rate of the error probability for codes on DSPCs with feedback. Proof of Theorem 3 rely on the averaging technique of Augustin [8], [9], Taylor’s expansion idea of Sheverdyaev [48] and the dummy channel method of Haroutunian [28], [29]. Our proof is substantially different from the proofs suggested by Sheverdyaev [48] and Augustin [9], previously. In addition, Lemma 15 and Lemma 16 are new to the best of our knowledge. Stationarity of the channel and finiteness of \mathcal{W} and \mathcal{Y} are implicit hypotheses for DSPCs with feedback. We can not remove the stationarity assumption altogether, within the confines of our approach, unless we strengthen our preliminary results significantly. Nevertheless, Theorem 3 applies to any sequence of DSPCs with feedback satisfying⁶ Assumption 4. The finiteness assumption on \mathcal{W} and \mathcal{Y} , on the other hand, is an artifact of our reluctance to focus on certain measurability issues. We believe Theorem 3 applies to any sequence of product channels with feedback satisfying Assumption 4 for a \mathcal{W}_0 such that $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}_0} = 0$.
- (IV) Theorem 4 establishes the sphere packing bound for Poisson channels $\Lambda^{[T,a,b,\varrho]}$, $\Lambda^{[T,a,b,\leq\varrho]}$, $\Lambda^{[T,a,b,\geq\varrho]}$, and $\Lambda^{[T,a,b]}$ with a polynomial prefactor, i.e. with a prefactor of the form $e^{-O(\ln T)}$. Wyner’s original result in [55],[56], for $\Lambda^{[T,a,b,\leq\varrho]}$ and $\Lambda^{[T,a,b]}$, have an exponential prefactor, i.e. a prefactor of the form $e^{-o(T)}$. Later, Burnashev and Kutoyants [12] improved Wyner’s result for the zero dark current case, i.e. for $\Lambda^{[T,0,b,\leq\varrho]}$ and $\Lambda^{[T,0,b]}$.
- (V) In Appendix A, we introduce the concept of minimum σ -algebra for information transmission and relate it to the concept of transition probability. Lemma 18 and Lemma 19 are new to the best of our knowledge, but they are stating observations about fundamental concepts, which are easy to verify and, in a sense, expected. Thus Lemmas 18 and 19 might have been reported in a different form, previously.

B. The Channel Coding Problem

A channel code is a strategy to convey from the transmitter at the input of the channel to the receiver at the output of the channel, a random choice from a finite message set. Once the transmitter and receiver agree on a strategy, the transmitter is given an element from the message set. Then the transmitter chooses the channel input, according to the strategy using the message. The channel input determines the probabilistic behavior of the channel output. The receiver observes the realization of the channel output and then chooses the decoded list based on the channel output, according to the strategy. If the message given to the transmitter is in the decoded list determined by the receiver then the transmission is successful, else an error is said to occur. Let us proceed with the formal definitions of these concepts.

Definition 1. A *channel* $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is an ordered pair composed of a measurable space $(\mathcal{Y}, \mathcal{Y})$, called the output space, and a set of probability measures \mathcal{W} on $(\mathcal{Y}, \mathcal{Y})$, called the input set. \mathcal{Y} is called the output set and \mathcal{Y} is called the σ -algebra of the output events. A channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is called a *discrete channel* if both \mathcal{Y} and \mathcal{W} are finite sets.

We denote the channel simply by \mathcal{W} rather than $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ whenever the output space is clear from the context. For the purposes of the channel coding problem Definition 1 suffices. However, while analyzing other information transmission problems —such as joint source channel coding problem— one is compelled to introduce a σ -field on \mathcal{W} and work with the transition probabilities, see Appendix A for a more detailed discussion.

⁵Definitions of the Poisson channels $\Lambda^{[T,a,b,\varrho]}$, $\Lambda^{[T,a,b,\leq\varrho]}$, $\Lambda^{[T,a,b,\geq\varrho]}$, and $\Lambda^{[T,a,b]}$ can be found in [35, Examples 8, 9, 10]

⁶It seems certain claims of Augustin in [9] on the necessity of stationarity for establishing sphere packing bounds are not accurate.

Definition 2. An (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is an ordered pair (Ψ, Θ) composed of an *encoding function* Ψ and a *decoding function* Θ :

- An *encoding function* is a function from the message set $\mathcal{M} \triangleq \{1, 2, \dots, M\}$ to the input set \mathcal{W} .
- A *decoding function* is a measurable function from the output space $(\mathcal{Y}, \mathcal{Y})$ to the set⁷ $\widehat{\mathcal{M}} \triangleq \{\mathcal{L} : \mathcal{L} \subset \mathcal{M} \text{ and } |\mathcal{L}| = L\}$.

Error event depends not only on the channel output but also on the transmitted message; hence unlike the decoded list, the error event is not measurable in the output space. But for each member of the message set one can calculate the conditional error probability.

Definition 3. Given an (M, L) channel code (Ψ, Θ) on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$, for each $m \in \mathcal{M}$ the *conditional error probability* P_e^m is

$$P_e^m \triangleq \int \mathbf{1}_{\{m \notin \widehat{\mathcal{M}}\}} w(dy) \quad \text{where } \Psi(m) = w \text{ and } \Theta(y) = \widehat{m}. \quad (1)$$

The *average error probability* P_e^{av} and the *maximum error probability* P_e^{\max} are

$$P_e^{av} \triangleq \frac{1}{M} \sum_{m \in \mathcal{M}} P_e^m \quad P_e^{\max} \triangleq \bigvee_{m \in \mathcal{M}} P_e^m. \quad (2)$$

For a given channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ the triplet (M, L, P_e) is achievable if there exists an (M, L) channel code with error probability less than or equal to P_e . Broadly speaking point-to-point channel coding problem aims to characterize achievable (M, L, P_e) triplets. The abstract formulation given above is general enough to cover almost all point-to-point channel coding problems as a special case. With the same token, however, it has scant structure to establish inner and outer bounds to the performance that are provably close to one another. The product structure, discussed in the following section, is commonly assumed in order establish such inner and outer bounds.

C. Product Channels, Stationarity and Memorylessness

Definition 4. A channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is a *product channel for finite index set* \mathcal{T} , if there exist channels $((\mathcal{Y}_t, \mathcal{Y}_t), \mathcal{W}_t)$ for $t \in \mathcal{T}$ such that $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$, $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$ and $\mathcal{W} = \mathcal{W}_{\mathcal{T}}$ where⁸

$$\mathcal{Y}_{\mathcal{T}} \triangleq \prod_{t \in \mathcal{T}}^{\times} \mathcal{Y}_t \quad \mathcal{Y}_{\mathcal{T}} \triangleq \prod_{t \in \mathcal{T}}^{\otimes} \mathcal{Y}_t \quad \mathcal{W}_{\mathcal{T}} \triangleq \left\{ w : w = \prod_{t \in \mathcal{T}}^{\otimes} w_t : w_t \in \mathcal{W}_t \right\}. \quad (3)$$

A product channel is *stationary* iff all $((\mathcal{Y}_t, \mathcal{Y}_t), \mathcal{W}_t)$ are identical.

Definition 5. A channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is a *memoryless channel for finite index set* \mathcal{T} , if there exists a product channel $((\mathcal{Y}_{\mathcal{T}}, \mathcal{Y}_{\mathcal{T}}), \mathcal{W}_{\mathcal{T}})$ such that $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$, $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$ and $\mathcal{W} \subset \mathcal{W}_{\mathcal{T}}$.

For any discrete channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$, n ‘independent’ uses of it given by $((\mathcal{Y}_{[1,n]}, \mathcal{Y}_{[1,n]}), \mathcal{W}_{[1,n]})$ is not only a memoryless channel, but also a stationary product channel. We call these channels discrete stationary product channels (DSPCs). They are customarily called discrete memoryless channels (DMCs).

For any $T \in \mathbb{R}^+$ the family of Poisson point processes on $(0, T]$ with bounded intensity functions $\Lambda^{[T,a,b]}$ described in⁹ [35, Definition 10] is a stationary product channel for any finite index set \mathcal{T} for $\mathcal{W}_t = \Lambda^{[\frac{T}{|\mathcal{T}|}, a, b]}$ for all $t \in \mathcal{T}$. Other families of Poisson processes described in [35, Definition 10], i.e. $\Lambda^{[T,a,b,\varrho]}$, $\Lambda^{[T,a,b,<\varrho]}$ and $\Lambda^{[T,a,b,>\varrho]}$, are not product channels, but they are memoryless channels.

A well known fact about DSPCs is that their Renyi capacities are additive, see [24, Theorem 5], [25, p149-150, eq 5.6.59]. Similarly, Renyi capacities are additive for $\Lambda^{[T,a,b]}$ considered in [35, Example 10]. In fact, the additivity of Renyi capacities holds for arbitrary product channels.

Lemma 1. Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a product channel for finite index set \mathcal{T} . Then

$$C_{\alpha, \mathcal{W}} = \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t} \quad \forall \alpha \in (0, \infty]. \quad (4)$$

Furthermore, if $C_{\alpha, \mathcal{W}} < \infty$ for an $\alpha \in (0, \infty]$ then $q_{\alpha, \mathcal{W}} = \prod_{t \in \mathcal{T}}^{\otimes} q_{\alpha, \mathcal{W}_t}$.

Proof of Lemma 1: Note that Lemma 1 is nothing but a restatement of [35, Lemma 19]. ■

We have introduced and discussed product channels and memoryless channels for finite index sets. Such an approach is general enough to capture these concepts completely for discrete time systems. For continuous time systems although finite index sets are good enough for some applications, a more complete picture can be obtained only by considering infinite index sets and Kolmogorov extension theorem [20, 12.12].

⁷Measurability of a function depend on the σ -algebra of the range space. We are using the σ -algebra in which all subsets of $\widehat{\mathcal{M}}$ are measurable, i.e. $\mathcal{Z}^{\widehat{\mathcal{M}}}$. For infinite sets such a choice is usually problematic. That is not the case here, because the range space $\widehat{\mathcal{M}}$ is a finite set.

⁸The existence and the uniqueness of product measures are guaranteed for any finite collection of σ -finite measures by [20, Theorem 4.4.4] and for any countable collections of probability measures by [20, Theorem 8.2.2].

⁹In [35, Section V-C], we have discussed Poisson point processes on the closed interval $[0, T]$, instead of the half open interval $(0, T]$. But our discussion holds pretty much as is for Poisson point processes on the half open interval $(0, T]$.

D. Product Channels With Feedback

In a product channel one has two distinct product structures: the product structure of the output space and the product structure of the members of the input set, i.e. the independence of the observations for the component channels for each member of the input set. In a product channel with feedback, the product structure of the output space is retained, but the product structure of the members of the input set is replaced with a dependence structure that can be described via transition probabilities. This dependence structure is not as straightforward as the independence, i.e. product of probability measures. We start with a brief discussion of the construction of probability measures using transition probabilities.

Let $(\mathcal{Y}_t, \mathcal{Y}_t)$ be a measurable space for each $t \in \mathbb{Z}^+$ and w_1 be a probability measure on $(\mathcal{Y}_1, \mathcal{Y}_1)$. Furthermore, let w_t be a transition probability from $(\mathcal{Y}_{(0,t)}, \mathcal{Y}_{(0,t)})$ to $(\mathcal{Y}_t, \mathcal{Y}_t)$ for each $t \in \mathbb{Z}^+ \setminus \{1\}$ where¹⁰ $\mathcal{Y}_{(0,t)} = \prod_{j \in \{1, \dots, t-1\}} \mathcal{Y}_j$ and $\mathcal{Y}_{(0,t)} = \prod_{j \in \{1, \dots, t-1\}}^\infty \mathcal{Y}_j$. By [11, Theorem 10.7.2] for any pair of measurable spaces $(\mathcal{Y}_1, \mathcal{Y}_1)$ and $(\mathcal{Y}_2, \mathcal{Y}_2)$, any measure w_1 on $(\mathcal{Y}_1, \mathcal{Y}_1)$ and any transition probability w_2 from $(\mathcal{Y}_1, \mathcal{Y}_1)$ to $(\mathcal{Y}_2, \mathcal{Y}_2)$, there exists a unique probability measure $w_{(0,2]}$ on $(\mathcal{Y}_{(0,2]}, \mathcal{Y}_{(0,2]})$ such that

$$w_{(0,2]}(\mathcal{E}_1 \times \mathcal{E}_2) \triangleq \int_{\mathcal{E}_1} \int_{\mathcal{E}_2} w_2(y_{(0,2)} | dy_2) w_1(dy_1) \quad \forall \mathcal{E}_1 \in \mathcal{Y}_1, \mathcal{E}_2 \in \mathcal{Y}_2.$$

Furthermore, $\int f w_{(0,2]}(dy_{(0,2]}) = \int \int f w_2(y_{(0,2)} | dy_2) w_1(dy_1)$ for any $f \in \mathcal{L}_1(w_{(0,2]})$ and for any non-negative measurable function f on $(\mathcal{Y}_{(0,2]}, \mathcal{Y}_{(0,2]})$. By repeating the same argument recursively, we can conclude that for any positive integer n there exists a unique probability measure $w_{(0,n]}$ on $(\mathcal{Y}_{(0,n]}, \mathcal{Y}_{(0,n]})$ such that,

$$w_{(0,n]} \left(\prod_{t \in (0,n]}^\times \mathcal{E}_t \right) \triangleq \int_{\mathcal{E}_1} \int_{\mathcal{E}_2} \dots \int_{\mathcal{E}_n} w_n(y_{(0,n)} | dy_n) \dots w_2(y_{(0,2)} | dy_2) w_1(dy_1) \quad \forall \mathcal{E}_1 \in \mathcal{Y}_1, \dots, \forall \mathcal{E}_n \in \mathcal{Y}_n. \quad (5)$$

Furthermore, $\int f w_{(0,n]}(dy_{(0,n]}) = \int \int \dots \int f w_n(y_{(0,n)} | dy_n) \dots w_2(y_{(0,2)} | dy_2) w_1(dy_1)$ for any $f \in \mathcal{L}_1(w_{(0,n]})$ and for any non-negative measurable function f on $(\mathcal{Y}_{(0,n]}, \mathcal{Y}_{(0,n]})$. Note that for $j, i \in \mathbb{Z}^+$ such that $j < i$, $w_{(0,j]}$ is the $(\mathcal{Y}_{(0,j]}, \mathcal{Y}_{(0,j]})$ marginal of $w_{(0,i]}$, i.e. $w_{(0,j]}(\mathcal{E}) = w_{(0,i]}(\mathcal{E} \times \mathcal{Y}_{(j,i]})$ for any $\mathcal{E} \in \mathcal{Y}_{(0,j]}$, by construction. This construction works only when n is finite; and that is good enough for our purposes. But it is also known —because of a result by C. T. Ionescu Tulcea [11, Theorem 10.7.3]— that there exists a unique probability measure $w_{(0,\infty]}$ on $(\prod_{t \in \mathbb{Z}^+}^\times \mathcal{Y}_t, \prod_{t \in \mathbb{Z}^+}^\infty \mathcal{Y}_t)$ satisfying above mentioned consistency property.

In order to assert the existence of a probability measure satisfying equation (5) we have assumed w_t 's to be transition probabilities. For w_t to be a transition probability two separate constraints have to be satisfied.

- (i) For all $y_{(0,t)} \in \mathcal{Y}_{(0,t)}$, $w_t(y_{(0,t)} | \cdot)$ is a probability measure on $(\mathcal{Y}_t, \mathcal{Y}_t)$.
- (ii) For all $\mathcal{E} \in \mathcal{Y}_t$, $w_t(\cdot | \mathcal{E}) : \mathcal{Y}_{(0,t)} \rightarrow [0, 1]$ is a $(\mathcal{Y}_{(0,t)}, \mathcal{B}([0, 1]))$ -measurable function.

One might suspect, at first, that the second constraint is superficial for asserting the existence of $w_{(0,n]}$ given in equation (5). This, however, is not true: the existence of $w_{(0,n]}$ does not follow from the first constraint, unless one qualifies w_t 's further. Consider the case $n = 2$ and $\mathcal{Y}_1 \neq \mathcal{Y}_1$. Then there exists a $\mathcal{E}_1 \subset \mathcal{Y}_1$ such that $\mathcal{E}_1 \notin \mathcal{Y}_1$. Let $w_2 = \mathbb{1}_{\{y_1 \in \mathcal{E}_1\}} \hat{w} + \mathbb{1}_{\{y_1 \notin \mathcal{E}_1\}} \tilde{w}$. Then the first constraint is satisfied. But there is no measure satisfying equation (5), unless¹¹ $w_1(\mathcal{E}_1) = 0$ or $w_1(\mathcal{E}_1) = 1$.

If $\mathcal{Y}_t = \mathcal{Y}_t$ for $t \in \mathbb{Z}^+$, every function on $\mathcal{Y}_{(0,t)}$ is measurable because all subsets of $\mathcal{Y}_{(0,t)}$ are in $\mathcal{Y}_{(0,t)}$. Thus the second constraint for being a transition probability is always satisfied. This is the reason why DSPCs with feedback can be introduced and analyzed without discussing the measurability issues at all. When \mathcal{Y} is not a finite set, however, \mathcal{Y} is rarely —if ever— the power set of \mathcal{Y} .

Definition 6. A channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is said to be a *product channel with feedback for finite ordered index set* $\mathcal{T} = \{1, \dots, n\}$ if there exist channels $((\mathcal{Y}_t, \mathcal{Y}_t), \mathcal{W}_t)$ for all $t \in \mathcal{T}$ such that $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$, $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$ and $\mathcal{W} = \mathcal{W}_{\vec{\mathcal{T}}}$ where

$$\mathcal{W}_{\vec{\mathcal{T}}} \triangleq \{w_{(0,n]} : w_t(y_{(0,t)} | \cdot) \in \mathcal{W}_t \quad \forall y_{(0,t)} \in \mathcal{Y}_{(0,t)} \text{ and } w_t \in \mathcal{P}((\mathcal{Y}_{(0,t)}, \mathcal{Y}_{(0,t)}) | (\mathcal{Y}_t, \mathcal{Y}_t)) \text{ for all } t \in \mathcal{T}\}. \quad (6)$$

A product channel with feedback is *stationary* iff all $((\mathcal{Y}_t, \mathcal{Y}_t), \mathcal{W}_t)$'s are identical.

Definition 7. A channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ is said to be a *memoryless channel with feedback for finite ordered index set* $\mathcal{T} = \{1, \dots, n\}$ if there exist channels $((\mathcal{Y}_t, \mathcal{Y}_t), \mathcal{W}_t)$ for $t \in \mathcal{T}$ such that $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$, $\mathcal{Y} = \mathcal{Y}_{\mathcal{T}}$ and $\mathcal{W} \subset \mathcal{W}_{\vec{\mathcal{T}}}$.

We called above family of channels memoryless channels with feedback, in order to be consistent with the convention that is already in use, i.e. "DMCs with feedback." However, considering the probabilistic behavior described, the word memoryless is misleading. Product space channels with feedback seems to be more accurate and descriptive, if somewhat mouthful.

Equation (6) describes the input set of product channels with feedback via the transition probabilities. For each transition probability w_t there is a corresponding function $\Psi_t : \mathcal{Y}_{(0,t)} \rightarrow \mathcal{W}_t$. These functions give an alternative description of $\mathcal{W}_{\vec{\mathcal{T}}}$ that is equivalent to the one given in equation (6). DSPCs with feedback are often described in terms of these functions, rather than the transition probabilities.

¹⁰Note that $(0, t-1] = (0, t)$ for any $t \in \mathbb{Z}^+$.

¹¹Let us consider the event $\mathcal{Y}_1 \times \mathcal{E}_2$ for some $\mathcal{E}_2 \in \mathcal{Y}_2$. Note that $\mathcal{Y}_1 \times \mathcal{E}_2 \in \mathcal{Y}_{\mathcal{T}}$ but $\{w_2(\cdot | \mathcal{E}_2) = \hat{w}(\mathcal{E}_2)\}$ is not a measurable event in $\mathcal{Y}_{\mathcal{T}}$ by construction.

In order to avoid discussing the measurability issues explicitly one might consider describing product channels with feedback, through the conditional probabilities.¹² For that approach the measurability constraints emerge in the form of existence of the regular conditional probabilities. If we want to handle the situation in the generality we are handling now, there is, practically, no difference between these two approaches. However, with appropriate assumptions the regular conditional probabilities are guaranteed to exist, and one can avoid discussing the measurability issues explicitly. For example, if the output σ -algebra of each component channel is the Borel σ -algebra of a Polish space then the same structure exists for the output σ -algebra of the channel and the conditional probabilities are guaranteed to exist by [20, Theorem 10.2.2].

Product channels and memoryless channels are special cases of memoryless channels with feedback because $\mathcal{W}_{\mathcal{T}} \subset \mathcal{W}_{\vec{\mathcal{T}}}$. An immediate consequence of the observation $\mathcal{W}_{\mathcal{T}} \subset \mathcal{W}_{\vec{\mathcal{T}}}$ is that $C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}} \geq C_{\alpha, \mathcal{W}_{\mathcal{T}}}$. More interestingly, the reverse inequality $C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}} \leq C_{\alpha, \mathcal{W}_{\mathcal{T}}}$ is true, as well.

Lemma 2. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a product channel with feedback for finite ordered index set $\mathcal{T} = \{1, 2, \dots, n\}$ then*

$$C_{\alpha, \mathcal{W}} = \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t} \quad \forall \alpha \in (0, \infty]. \quad (7)$$

Furthermore, if $C_{\alpha, \mathcal{W}} < \infty$ for an $\alpha \in (0, \infty]$ then $q_{\alpha, \mathcal{W}} = \prod_{t \in \mathcal{T}}^{\otimes} q_{\alpha, \mathcal{W}_t}$.

For the case when component channels are discrete channels Lemma 2 has been common knowledge among the researcher working on error exponents with feedback for some time now. Augustin [9, p.304-306] mentions an equivalent claim¹³ without proof for the case when input sets, i.e. \mathcal{W}_t 's, are finite sets.

Proof of Lemma 2: We prove the lemma for $\alpha \in \mathbb{R}^+ \setminus \mathbb{Z}^+$ in the following. This implies $C_{\alpha, \mathcal{W}} = \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t}$ for all $\alpha \in (0, \infty]$, because the Renyi capacity is an increasing lower semicontinuous function of the order by [35, Lemma 11-(a)]. The claim about the Renyi centers follows from the corresponding claim in Lemma 1 and the uniqueness of the Renyi centers, i.e. [35, Theorem 1].¹⁴

Any probability measure on $(\mathcal{Y}_t, \mathcal{Y}_t)$ is a transition probability from $(\mathcal{Y}_{(0,t)}, \mathcal{Y}_{(0,t)})$ to $(\mathcal{Y}_t, \mathcal{Y}_t)$ by definition. Thus $\mathcal{W}_{\mathcal{T}}$ is a subset of $\mathcal{W}_{\vec{\mathcal{T}}}$, i.e. $\mathcal{W}_{\mathcal{T}} \subset \mathcal{W}_{\vec{\mathcal{T}}}$, and $C_{\alpha, \mathcal{W}_{\mathcal{T}}} \leq C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}}$. Then $\sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t} \leq C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}}$ by Lemma 1 and equation (7) holds if $C_{\alpha, \mathcal{W}_t}$ is infinite for a $t \in \mathcal{T}$. Thus, we assume that $C_{\alpha, \mathcal{W}_t} < \infty$ for all $t \in \mathcal{T}$ for the rest of the proof. Then for each $t \in \mathcal{T}$, \mathcal{W}_t has a unique Renyi center $q_{\alpha, \mathcal{W}_t}$ by [35, Theorem 1]. Let q be the product measure for $q_{\alpha, \mathcal{W}_t}$'s, i.e.

$$q = \prod_{t \in \mathcal{T}}^{\otimes} q_{\alpha, \mathcal{W}_t}.$$

We show in the following that

$$D_{\alpha}(w \| q) \leq \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t} \quad \forall w \in \mathcal{W}_{\vec{\mathcal{T}}}. \quad (8)$$

Then $C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}} \leq \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t}$ by [35, Lemma 13]. Then $C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}} = \sum_{t \in \mathcal{T}} C_{\alpha, \mathcal{W}_t}$ and $q = q_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}}$ by [35, Theorem 1].

For any $w \in \mathcal{W}_{\vec{\mathcal{T}}}$ there exist transition probabilities $w_t \in \mathcal{P}((\mathcal{Y}_{(0,t)}, \mathcal{Y}_{(0,t)}) | (\mathcal{Y}_t, \mathcal{Y}_t))$ by Definition 6. Let us define another sequence of transition probabilities as follows.

$$\nu_t \triangleq \frac{1}{2} w_t + \frac{1}{2} q_{\alpha, \mathcal{W}_t} \quad \forall t \in \mathcal{T}.$$

Using ν_t 's and [11, Theorem 10.7.2], we define the probability measure ν on $(\mathcal{Y}, \mathcal{Y})$. This construction implies not only the absolute continuity of w and q in ν , but also $\frac{dw}{d\nu} \leq 2^n$ and $\frac{dq}{d\nu} \leq 2^n$. Furthermore,

$$\frac{dw}{d\nu} = \prod_{t=1}^n \frac{dw_t}{d\nu_t} \quad \frac{dq}{d\nu} = \prod_{t=1}^n \frac{dq_{\alpha, \mathcal{W}_t}}{d\nu_t} \quad \forall y \in \mathcal{Y}$$

where $\frac{dw_t}{d\nu_t}$ and $\frac{dq_{\alpha, \mathcal{W}_t}}{d\nu_t}$ are $(\mathcal{Y}_{(0,t)}, \mathcal{B}(\mathbb{R}))$ -measurable functions. Then by [11, Theorem 10.7.2] we have¹⁵

$$\begin{aligned} \int \left[\frac{dw}{d\nu} \right]^{\alpha} \left[\frac{dq}{d\nu} \right]^{1-\alpha} \nu(dy) &= \int \int \dots \int \left[\frac{dw}{d\nu} \right]^{\alpha} \left[\frac{dq}{d\nu} \right]^{1-\alpha} \nu_n(y_{(0,n)} | dy_n) \dots \nu_2(y_{(0,2)} | dy_2) \nu_1(dy_1) \\ &= \int \int \dots \prod_{t=1}^{n-1} \left[\frac{dw_t}{d\nu_t} \right]^{\alpha} \left[\frac{dq_{\alpha, \mathcal{W}_t}}{d\nu_t} \right]^{1-\alpha} \int \left[\frac{dw_n}{d\nu_n} \right]^{\alpha} \left[\frac{dq_{\alpha, \mathcal{W}_n}}{d\nu_n} \right]^{1-\alpha} \nu_n(y_{(0,n)} | dy_n) \dots \nu_2(y_{(0,2)} | dy_2) \nu_1(dy_1) \end{aligned} \quad (9)$$

On the other hand $w_t(y_{(0,t)} | \cdot) \in \mathcal{W}_t$ for all $y_{(0,t)}$ by Definition 6. Thus, as a result of [35, Theorem 1] we have

$$\frac{1}{\alpha-1} \ln \int \left[\frac{dw_t}{d\nu_t} \right]^{\alpha} \left[\frac{dq_{\alpha, \mathcal{W}_t}}{d\nu_t} \right]^{1-\alpha} \nu_t(y_{(0,t)} | dy_t) \leq C_{\alpha, \mathcal{W}_t} \quad \forall y_{(0,t)} \in \mathcal{Y}_{(0,t)}. \quad (10)$$

Equation (8) follows from equations (9) and (10). ■

¹²A concise exposition of the concept of regular conditional probabilities can be found in [20, Section 10.2].

¹³ $\exp(\frac{\alpha-1}{\alpha} C_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}}) q_{\alpha, \mathcal{W}_{\vec{\mathcal{T}}}} = \exp(\frac{\alpha-1}{\alpha} C_{\alpha, \mathcal{W}_{\mathcal{T}}}) q_{\alpha, \mathcal{W}_{\mathcal{T}}}$ whenever $q_{\alpha, \mathcal{W}_{\mathcal{T}}}$ is defined.

¹⁴In fact, proving Lemma 2 for any dense subset of \mathbb{R}^+ is sufficient.

¹⁵[11, Theorem 10.7.2] confines its claims to integrable functions, but the equality of the integrals holds for all non-negative functions, as well.

II. OPERATIONAL SIGNIFICANCE OF THE RENYI CAPACITY AND ASYMPTOTIC QUANTITIES

The information transmission problems did not play any role in the definition or in the analysis of the Renyi capacity presented in the first paper of the series [35]. In information theorists' parlance: Renyi capacity does not have any operational significance because of its definition. Channel coding theorems and their converses establish the operational significance of the Renyi capacity. In this section we review two well known results that quantify this operational significance.

For any DSPC, the order one Renyi capacity is equal to the channel capacity [46], i.e. for any DSPC the order one Renyi capacity is equal to the threshold for the rates below which reliable communication is possible and above which reliable communication is impossible. Renyi capacities of other orders bound the performance of channel codes through the sphere packing exponent [7, Theorem 1], [24, Theorem 1].

We define the sphere packing exponent and review some of its properties as a function of the rate in Section II-A. We derive Gallager's inner bound in Section II-B and Arimoto's outer bound in Section II-C. Then in Section II-D we define the concept of channel capacity formally and demonstrate that Gallager's inner bound and Arimoto's outer bound determine the channel capacity for a class of sequence of channels much broader than the DSPCs.

A. The Sphere Packing Exponent

Definition 8. For any channel $((\mathcal{Y}, \mathcal{V}), \mathcal{W})$ and rate $R \in \mathbb{R}^+$ the sphere packing exponent is¹⁶

$$E_{sp}(R, \mathcal{W}) \triangleq \sup_{\alpha \in (0, \infty)} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) \quad \forall R \in \mathbb{R}^+. \quad (11)$$

Using the continuity and the monotonicity of $C_{\alpha, \mathcal{W}}$ in α , we can obtain an alternative expression for $E_{sp}(R, \mathcal{W})$.

Lemma 3. For any channel, $E_{sp}(R, \mathcal{W})$ is convex in R on \mathbb{R}^+ , finite on $(\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, \infty)$ and continuous on the closure of $(\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, \infty)$ in \mathbb{R}^+ . Furthermore, $E_{sp}(R, \mathcal{W})$ is decreasing on $(0, C_{1, \mathcal{W}}]$ and increasing on $[C_{1, \mathcal{W}}, \infty)$. In particular,

$$E_{sp}(R, \mathcal{W}) = \begin{cases} \infty & R < \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} \\ \sup_{\alpha \in (0, 1)} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) & R = \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} \\ \sup_{\alpha \in [\phi, 1)} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) & R = C_{\phi, \mathcal{W}} \text{ for some } \phi \in (0, 1) \\ 0 & R = C_{1, \mathcal{W}} \\ 0 & R \geq C_{\chi, \mathcal{W}} \text{ and } \chi = 1 \\ \sup_{\alpha \in [1, \phi]} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) & R = C_{\phi, \mathcal{W}} \text{ for some } \phi \in (1, \chi) \\ \sup_{\alpha \in [1, \chi)} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) & R \geq C_{\chi, \mathcal{W}} \text{ and } \chi > 1 \end{cases} \quad \text{where } \chi = \sup\{\alpha : C_{\alpha, \mathcal{W}} < \infty\}. \quad (12)$$

Proof of Lemma 3: $E_{sp}(R, \mathcal{W})$ is convex in R , because the pointwise supremum of a family of convex functions is convex and $\frac{\alpha-1}{\alpha}(C_{\alpha, \mathcal{W}} - R)$ is convex in R or any $\alpha \in \mathbb{R}^+$. Monotonicity claims follow from equation (12) established in the following. The finiteness and the continuity claims are proved while establishing equation (12).

Recall that $C_{\alpha, \mathcal{W}}$ is an increasing function of the order α by [35, Lemma 11-(a)].

- If $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} = \infty$ then $C_{1/2, \mathcal{W}} = \infty$ and $E_{sp}(R, \mathcal{W}) = \infty$ for all $R \in \mathbb{R}^+$. On the other hand $R < \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}$ for all $R \in \mathbb{R}^+$. Hence (12) holds.

Claims about the continuity and the finiteness of $E_{sp}(R, \mathcal{W})$ are void.

- If $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} < \infty$ and $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} = C_{1, \mathcal{W}}$ then $E_{sp}(R, \mathcal{W}) = \infty$ for all $R \in (0, C_{1, \mathcal{W}})$ and $E_{sp}(C_{1, \mathcal{W}}, \mathcal{W}) = 0$. For $R > C_{1, \mathcal{W}}$, expression $\frac{1-\alpha}{\alpha}(C_{\alpha, \mathcal{W}} - R)$ is non-negative only for $\alpha \geq 1$ satisfying $C_{\alpha, \mathcal{W}} \leq R$. Thus equation (12) holds.¹⁷ If $\chi = 1$ then $E_{sp}(R, \mathcal{W}) = 0$ for all $R > C_{1, \mathcal{W}}$ and $E_{sp}(R, \mathcal{W})$ is finite and continuous on $(\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, \infty)$.

If $\chi > 1$ then depending on the finiteness of $C_{\chi, \mathcal{W}}$ either one or both of the last two cases happen in equation (12). For both $C_{\chi, \mathcal{W}} = \infty$ case and $C_{\chi, \mathcal{W}} < \infty$ case, $E_{sp}(R, \mathcal{W}) \leq R - C_{1, \mathcal{W}} < \infty$. Thus $E_{sp}(R, \mathcal{W})$ is continuous from the right at $R = C_{1, \mathcal{W}}$. Furthermore, $E_{sp}(R, \mathcal{W})$ is continuous on $(C_{1, \mathcal{W}}, \infty)$ because convex functions are continuous on the interior of the interval they are finite, [20, Theorem 6.3.3].

- If $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} < \infty$ and $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} \neq C_{1, \mathcal{W}}$ then $E_{sp}(R, \mathcal{W}) = \infty$ for all $R \in (0, \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}})$. For $R \geq \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}$, the non-negativity of $\frac{1-\alpha}{\alpha}(C_{\alpha, \mathcal{W}} - R)$ imply the restrictions given in equation (12) for different intervals.

As a result of (12), $E_{sp}(R, \mathcal{W})$ is finite for all $R > \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}$. Thus, $E_{sp}(R, \mathcal{W})$ is continuous on $(\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}, \infty)$ by [20, Theorem 6.3.3]. In order to extend the continuity to the closure of $(\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}, \infty)$ in $(0, \infty)$, note that for each $\alpha \in (0, 1)$ the function $\frac{1-\alpha}{\alpha}(C_{\alpha, \mathcal{W}} - R)$ is decreasing and continuous in R . Thus $E_{sp}(R, \mathcal{W})$ is a decreasing and lower semicontinuous function of R on $(0, C_{1, \mathcal{W}})$. Hence $E_{sp}(R, \mathcal{W})$ is continuous from the right on $(0, C_{1, \mathcal{W}})$ and hence $E_{sp}(R, \mathcal{W})$ is continuous from the right at $R = \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}}$ if $\lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}} > 0$. ■

¹⁶In [36], Omura denotes the sphere packing exponent for the rates greater than the order one Renyi capacity by $C_{sp}(R)$.

¹⁷To be precise there is one case for which the restriction does not follow from the non-negativity of $\frac{1-\alpha}{\alpha}(C_{\alpha, \mathcal{W}} - R)$, alone. That is the case when both χ and $C_{\chi, \mathcal{W}}$ are finite, $\chi \neq 1$, and $R \geq C_{\chi, \mathcal{W}}$. In this case the non-negativity of $\frac{1-\alpha}{\alpha}(C_{\alpha, \mathcal{W}} - R)$ implies the restriction to be $[1, \chi]$ rather than $[1, \chi)$. But the value of the supremum on $[1, \chi]$ is equal to the value of the supremum on $[1, \chi)$ because $C_{\alpha, \mathcal{W}}$ is continuous from the left. Note that $C_{\alpha, \mathcal{W}}$ is continuous from the left because it is increasing and lower semicontinuous by [35, Lemma 11-(a)].

B. Gallager's Inner Bound

For rates less than $C_{1,\mathcal{W}}$, the sphere packing exponent confines the optimal performance for the channel coding problem via Gallager's inner bound, [24, Theorem 1], [25, Problem 5.20, p 538]. Let us start with deriving Gallager's inner bound for list decoding.¹⁸

Lemma 4. *For any channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$, prior $p \in \mathcal{P}(\mathcal{W})$, message set size $M \in \mathbb{Z}^+$, list size $L \in \{1, \dots, M-1\}$ and order $\alpha \in [\frac{1}{1+L}, 1)$ there exists an (M, L) channel code such that*

$$\ln P_e^{av} \leq \frac{\alpha-1}{\alpha} \left[I_\alpha(p; \mathcal{W}) - \frac{1}{L} \ln \binom{M-1}{L} \right]. \quad (13)$$

Furthermore, for any $\epsilon > 0$ and L, M satisfying $\ln \frac{eM}{L} \in [C_{\frac{1}{1+L}, \mathcal{W}}, C_{1, \mathcal{W}})$, there exists an (M, L) channel code such that

$$P_e^{av} \leq (1 + \epsilon) e^{-E_{sp}(\ln \frac{eM}{L}, \mathcal{W})}. \quad (14)$$

If $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}} = 0$ then equation (14) holds for $\epsilon = 0$ as well.¹⁹

Lemma is proved using a standard random coding argument for a maximum likelihood decoder. Gallager's inner bound is unique in the sense that it is proved without invoking probabilistic results, such as law of large numbers or central limit theorem. It is possible to strengthen the result by considering channels satisfying additional hypotheses. Altug and Wagner [2], [4] have shown that for DSPCs when $L = 1$ for large enough rates it is possible to replace the $(1 + \epsilon)$ term with an $O(n^{-\frac{1}{2\alpha}})$ term for certain $\alpha \in (1/2, 1)$, where n is the block length. Later Scarlett, Martinez and Fabregas [44] presented an alternative derivation of the result. First Scarlett, Martinez and Fabregas [45] and then Honda [32], derived approximations to random coding union bound which can be used to characterize the $O(n^{-\frac{1}{2\alpha}})$ term explicitly. Shannon Gallager and Berlekamp [47] proved that Gallager's bound is tight for DSPCs, in terms of the exponential decay rate of the error probability with the block length.

Proof of Lemma 4: Instead of bounding the error probability of a code, we bound the average error probability of an ensemble of codes. We assume that the messages are assigned to the members of \mathcal{W} , independently of one another and according to the p.m.f. $p \in \mathcal{P}(\mathcal{W})$. We use a maximum likelihood decoder: for each $y \in \mathcal{Y}$, decoded list $\Theta(y)$ is composed of L messages with the largest $\frac{d\psi(m)}{dq_{\alpha,p}}$ values. If there is a tie, decoder chooses the messages with the lower indices.

In order to bound the expected value of the average error probability of the code over the ensemble, let us consider the expected value of the conditional error probability of the message with the greatest index. An error will occur only when Radon-Nikodym derivative of L other messages is at least as large as the Radon-Nikodym derivative of the transmitted message. We bound this probability using an auxiliary threshold γ :

$$\mathbf{E}[P_e^{av}] \leq \sum_w p(w) \int_{\frac{dw}{dq_{\alpha,p}} \leq \gamma} \frac{dw}{dq_{\alpha,p}} q_{\alpha,p}(dy) + \binom{M-1}{L} \sum_w p(w) \int_{\frac{dw}{dq_{\alpha,p}} > \gamma} \left[\sum_{\tilde{w}} p(\tilde{w}) \mathbb{1}_{\{\frac{d\tilde{w}}{dq_{\alpha,p}} \geq \frac{dw}{dq_{\alpha,p}}\}} \right]^L \frac{dw}{dq_{\alpha,p}} q_{\alpha,p}(dy).$$

We bound the first and the second terms separately. Let us assume that $\frac{1}{1+L} \leq \alpha \leq 1$. Then

$$\begin{aligned} \int_{\frac{dw}{dq_{\alpha,p}} \leq \gamma} \frac{dw}{dq_{\alpha,p}} q_{\alpha,p}(dy) &\leq \gamma^{1-\alpha} \int_{\frac{dw}{dq_{\alpha,p}} \leq \gamma} \left(\frac{dw}{dq_{\alpha,p}} \right)^\alpha q_{\alpha,p}(dy) \\ \binom{M-1}{L} \int_{\frac{dw}{dq_{\alpha,p}} > \gamma} \left[\sum_{\tilde{w}} p(\tilde{w}) \mathbb{1}_{\{\frac{d\tilde{w}}{dq_{\alpha,p}} \geq \frac{dw}{dq_{\alpha,p}}\}} \right]^L \frac{dw}{dq_{\alpha,p}} q_{\alpha,p}(dy) &\leq \binom{M-1}{L} e^{(\alpha-1)I_\alpha(p; \mathcal{W})L} \int_{\frac{dw}{dq_{\alpha,p}} > \gamma} \left(\frac{dw}{dq_{\alpha,p}} \right)^{1-L\alpha} q_{\alpha,p}(dy) \\ &\leq \binom{M-1}{L} e^{(\alpha-1)I_\alpha(p; \mathcal{W})L} \gamma^{1-\alpha-\alpha L} \int_{\frac{dw}{dq_{\alpha,p}} > \gamma} \left(\frac{dw}{dq_{\alpha,p}} \right)^\alpha q_{\alpha,p}(dy) \end{aligned}$$

where $I_\alpha(p; \mathcal{W})$ is the order α Renyi information given in [35, Definition 4]. If we set $\gamma = \left[\binom{M-1}{L} \right]^{\frac{1}{L\alpha}} e^{\frac{\alpha-1}{\alpha} I_\alpha(p; \mathcal{W})}$ we get

$$\mathbf{E}[P_e^{av}] \leq \gamma^{1-\alpha} \sum_w p(w) \int \left(\frac{dw}{dq_{\alpha,p}} \right)^\alpha q_{\alpha,p}(dy) = \gamma^{1-\alpha} e^{(\alpha-1)I_\alpha(p; \mathcal{W})} = \left[\binom{M-1}{L} \right]^{\frac{1-\alpha}{L\alpha}} e^{\frac{\alpha-1}{\alpha} I_\alpha(p; \mathcal{W})}.$$

Since there exists a code with P_e^{av} less than or equal to $\mathbf{E}[P_e^{av}]$, there exists a code satisfying (13).

Using Stirling's approximation for factorials, i.e. $\sqrt{2\pi n} (n/e)^n \leq n! \leq e\sqrt{n} (n/e)^n$, and the identity $\ln x \leq x - 1$ we get

$$\frac{1}{L} \ln \binom{M-1}{L} \leq \frac{1}{L} \ln \frac{e\sqrt{M-1}}{2\pi\sqrt{L(M-1-L)}} + \ln \frac{M-1}{L} + \frac{M-1-L}{L} \ln \left(1 + \frac{L}{M-1-L} \right) \leq \ln \frac{M-1}{L} + 1.$$

¹⁸Lemma 4 is slightly different from [25, Problem 5.20, p538]. When we consider sequences of codes and deal with asymptotic quantities both results lead to the same conclusions provided that the list size L is bounded. However, if L is allowed to grow exponentially then only Lemma 4 lead to appropriate bounds that have matching converses.

¹⁹Recall that $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}} = 0$ iff there exists a $\mu \in \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ such that $\mathcal{W} \prec^{uni} \mu$ by [35, Lemma 21-(d)].

Then for any $\alpha \in [\frac{1}{1+L}, 1]$ we have

$$\ln P_e^{av} \leq \frac{1-\alpha}{\alpha} (\ln \frac{Me}{L} + \ln \frac{M-1}{M} - I_\alpha(p; \mathcal{W})).$$

In order to obtain (14), first note that for any $R \in [C_{\frac{1}{1+L}, \mathcal{W}}, C_{1, \mathcal{W}})$ and $\epsilon > 0$ there exists an $\alpha \in [\frac{1}{1+L}, 1)$ such that $\frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) > E_{sp}(R, \mathcal{W}) - \ln(1 + \epsilon)$ because of the alternative expression for $E_{sp}(R, \mathcal{W})$ given in equation (12). On the other hand, for any $M \geq 2$ and α there exists a prior $p \in \mathcal{P}(\mathcal{W})$ such that $I_\alpha(p; \mathcal{W}) \geq C_{\alpha, \mathcal{W}} - \ln \frac{M}{M-1}$ because $C_{\alpha, \mathcal{W}}$ is the supremum of $I_\alpha(p; \mathcal{W})$ over the priors. Thus, for any $\epsilon > 0$ and $L, M \in \mathbb{Z}^+$ such that $\ln \frac{eM}{L} \in [C_{\frac{1}{1+L}, \mathcal{W}}, C_{1, \mathcal{W}})$ there exists an $\alpha \in [\frac{1}{1+L}, 1)$ and $p \in \mathcal{P}(\mathcal{W})$ such that

$$E_{sp}(\frac{Me}{L}, \mathcal{W}) - \ln(1 + \epsilon) \leq \frac{1-\alpha}{\alpha} (I_\alpha(p; \mathcal{W}) - \ln \frac{Me}{L} - \ln \frac{M-1}{M}).$$

If $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}} = 0$ then $E_{sp}(R, \mathcal{W}) = \sup_{\alpha \in [\phi, 1]} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R)$ for R values satisfying $R = C_{\phi, \mathcal{W}}$ for a $\phi \in (0, 1)$, as a result of equation (12). Then as a result of the extreme value theorem [34, Theorem 27.4], for any $R \in [C_{\frac{1}{1+L}, \mathcal{W}}, C_{1, \mathcal{W}})$ there exists an $\alpha \in [\frac{1}{1+L}, 1]$ such that $\frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) = E_{sp}(R, \mathcal{W})$. Thus (14) holds for $\epsilon = 0$. ■

C. Arimoto's Outer Bound

For rates greater than $C_{1, \mathcal{W}}$ the sphere packing exponent confines the optimal performance for the channel coding problem through Arimoto's outer bound [7, Theorem 1].

Lemma 5. Any (M, L) channel code (Ψ, Θ) on a channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ satisfies

$$d_\alpha(P_e^{av} \| 1 - \frac{L}{M}) \leq I_\alpha(p; \mathcal{W}) \quad \forall \alpha \in \mathbb{R}^+ \quad (15)$$

where p is the probability mass function generated by the encoder Ψ on \mathcal{W} when each message has equal probability²⁰ and $d_\alpha(\cdot \| \cdot) : [0, 1] \times [0, 1] \rightarrow [0, \infty]$ is given by

$$d_\alpha(x \| z) \triangleq \begin{cases} \frac{1}{\alpha-1} \ln [x^\alpha z^{1-\alpha} + (1-x)^\alpha (1-z)^{1-\alpha}] & \alpha \in \mathbb{R}^+ \setminus \{1\} \\ x \ln \frac{x}{z} + (1-x) \ln \frac{1-x}{1-z} & \alpha = 1 \end{cases} \quad (16)$$

If $C_{1, \mathcal{W}} < \infty$ then any (M, L) channel code on \mathcal{W} such that $\ln \frac{M}{L} \geq C_{1, \mathcal{W}}$ satisfies

$$P_e^{av} \geq 1 - e^{-E_{sp}(\ln \frac{M}{L}, \mathcal{W})}. \quad (17)$$

First Omura [36] and then Dueck and Korner [21] have shown that Arimoto's outer bound is tight for DSPCs, in terms of the exponential decay rate of the probability of correct decoding with block length. If $C_{\alpha, \mathcal{W}} = \infty$ for all $\alpha > 1$ then $E_{sp}(R, \mathcal{W}) = 0$ for all $R \in [C_{1, \mathcal{W}}, \infty)$ and $E_{sp}(C_{1, \mathcal{W}}, \mathcal{W}) = 0$ for any \mathcal{W} such that $C_{1, \mathcal{W}} < \infty$. Thus, Arimoto's outer bound on the average error probability is non-trivial only when $\ln \frac{M}{L} > C_{1, \mathcal{W}}$ and $C_{\alpha, \mathcal{W}} < \infty$ for an $\alpha \in (1, \infty)$.

We derive Arimoto's outer bound given in (17) from the bound given in (15), which is a result of the monotonicity of Renyi divergence in the underlying σ -algebra, i.e. [35, Lemma 9-(f)], and the alternative expression for Renyi information given in [35, Lemma 10]. For $\alpha = 1$, the bound given in (15) is equivalent to Fano's inequality, [15, Theorem 3.8]. For $\alpha \in (0, 1)$ and $\alpha \in (1, \infty)$ monotonicity of Renyi divergence in the underlying σ -algebra is equivalent to the Holder's inequality and the reverse Holder's inequality, respectively. Sheverdyaev [48] is the first one to use reverse Holder's inequality to obtain a bound equivalent to (15) for $\alpha \in (1, \infty)$. Augustin obtained the same bound using convexity [9, p182]. More recently, Polyanskiy and Verdú [41] obtained a bound equivalent to (15) for $\alpha \in (0, 1) \cup (1, \infty)$.

Proof of Lemma 5: Let s be the probability measure on $(\mathcal{M} \times \mathcal{Y}, \mathcal{Z}^{\mathcal{M}} \otimes \mathcal{Y})$ generated by the code and the channel. Furthermore, let \tilde{s} be the product probability measure of the uniform distribution on $(\mathcal{M}, \mathcal{Z}^{\mathcal{M}})$ and $q_{\alpha, p}$ on $(\mathcal{Y}, \mathcal{Y})$. Then using the identity $I_\alpha(p; \mathcal{W}) = D_\alpha(p \circ \mathcal{W} \| p \times q_{\alpha, p})$ given in [35, equation (34) of Lemma 10] we get,

$$D_\alpha(s \| \tilde{s}) = I_\alpha(p; \mathcal{W}).$$

On the other hand, for any sub- σ -algebra \mathcal{A} of $\mathcal{Z}^{\mathcal{M}} \otimes \mathcal{Y}$ we have $D_\alpha^{\mathcal{A}}(s \| \tilde{s}) \leq D_\alpha(s \| \tilde{s})$ by [35, Lemma 9-(f)]. Then we get (15) by observing that $D_\alpha^{\mathcal{A}}(s \| \tilde{s}) = d_\alpha(P_e^{av} \| 1 - \frac{L}{M})$ if \mathcal{A} is the σ -algebra generated by the set $\{(m, y) : m \in \Theta(y)\}$.

Note that $d_\alpha(x \| z) \geq \frac{1}{\alpha-1} \ln(1-x)^\alpha (1-z)^{1-\alpha}$ for any $\alpha \in (1, \infty)$, $x \in [0, 1]$ and $z \in [0, 1]$. Then using the fact that $I_\alpha(p; \mathcal{W}) \leq C_{\alpha, \mathcal{W}}$ we get,

$$\frac{\alpha}{\alpha-1} \ln(1 - P_e^{av}) - \ln \frac{L}{M} \leq C_{\alpha, \mathcal{W}} \quad \forall \alpha \in (1, \infty) \quad \Rightarrow \quad P_e^{av} \geq 1 - e^{\frac{\alpha-1}{\alpha} (C_{\alpha, \mathcal{W}} - \ln \frac{M}{L})} \quad \forall \alpha \in (1, \infty).$$

Then equation (17) follows from the alternative expression for $E_{sp}(R, \mathcal{W})$ given (12). ■

²⁰Multiple messages can be assigned to the same member of \mathcal{W} by Definition 2. Thus p is not necessarily a uniform distribution on a size M subset of \mathcal{W} .

D. The Channel Capacity

Channel capacity is an asymptotic quantity that is defined for sequences of channels.

Definition 9. Given a sequence of channels $\{(\mathcal{Y}^{(n)}, \mathcal{Y}^{(n)}), \mathcal{W}^{(n)}\}_{n \in \mathbb{Z}^+}$ and a sequence of scaling factors $\{t^{(n)}\}_{n \in \mathbb{Z}^+}$, a sequence of codes $\{(\Psi^{(n)}, \Theta^{(n)})\}_{n \in \mathbb{Z}^+}$ is reliable iff

$$\lim_{n \rightarrow \infty} P_e^{av(n)} = 0. \quad (18)$$

A sequence of codes $\{(\Psi^{(n)}, \Theta^{(n)})\}_{n \in \mathbb{Z}^+}$ is of rate R iff

$$\liminf_{n \rightarrow \infty} \frac{1}{t^{(n)}} \ln \frac{M^{(n)}}{L^{(n)}} = R. \quad (19)$$

The channel capacity for the pair $\{(\mathcal{W}^{(n)}, t^{(n)})\}_{n \in \mathbb{Z}^+}$, denoted by $C(\mathcal{W}^{(n)}, t^{(n)})$, is the supremum of the rates of the reliable sequences of codes.

While discussing asymptotic quantities, such as channel capacity, we almost always consider sequences of channels that either satisfy the following two conditions, or can be expressed equivalently with a model satisfying them.

(C1) $(\mathcal{Y}^{(n)}, \mathcal{Y}^{(n)}) = (\mathcal{Y}, \mathcal{Y})$ for all $n \in \mathbb{Z}^+$

(C2) $\mathcal{W}^{(i)} \subset \mathcal{W}^{(j)}$ for all $i \leq j$

Consider, for example, the Poisson channels $\Lambda^{[T, a, b, \varrho]}$ described in [35, Example 8]. Let $\mathcal{W}^{(n)}$ be $\Lambda^{[n, a, b, \varrho]}$ and $t^{(n)}$ be n . When the problem is posed in this form, output space is different for each value of n . If we extend the intensity function in time horizon to infinity at fixed intensity level ϱ for each member of $\mathcal{W}^{(n)}$ for each n , we get a model that is equivalent to the original one. This equivalent model, however, satisfies the above mentioned conditions. This modification works for other Poisson channels described in [35, Examples 9, 10, and 11], as well. A similar modification works, if we are given a sequence of channels $\{\mathcal{W}_i\}_{i \in \mathbb{Z}^+}$ and $\mathcal{W}^{(n)}$ is the product channel $\mathcal{W}_{\mathcal{T}_n}$ —or the product channel with feedback $\mathcal{W}_{\vec{\mathcal{T}}_n}$ —for the index set $\mathcal{T}_n = \{1, \dots, n\}$.

The proof of the existence of reliable sequences for rates less than $C(\mathcal{W}^{(n)}, t^{(n)})$ is usually called the direct part. The proof of the non-existence of reliable sequences for rates greater than $C(\mathcal{W}^{(n)}, t^{(n)})$ is usually called the converse part. For certain sequences of channels, one can strengthen the converse part by proving that $P_e^{av(n)}$ converges to one for all rate R sequences of codes for any R greater than $C(\mathcal{W}^{(n)}, t^{(n)})$. These results are called the strong converses as opposed to the weak converses, which only establish that $P_e^{av(n)}$ is bounded away from zero.

Theorem 1. Let $\{(\mathcal{Y}^{(n)}, \mathcal{Y}^{(n)}), \mathcal{W}^{(n)}\}_{n \in \mathbb{Z}^+}$ be a sequence of channels and $\{t^{(n)}\}_{n \in \mathbb{Z}^+}$ be a sequence of scaling factors such that $\lim_{n \rightarrow \infty} t^{(n)} = \infty$.

(a) If there exists an $\varepsilon > 0$ and a lower semicontinuous function φ such that²¹

$$\lim_{n \rightarrow \infty} \frac{C_{\alpha, \mathcal{W}^{(n)}}}{t^{(n)}} = \varphi(\alpha) \quad \forall \alpha \in [1 - \varepsilon, 1] \quad (20)$$

then $C(\mathcal{W}^{(n)}, t^{(n)}) = \varphi(1)$.

(b) If there exists an $\varepsilon > 0$ and an upper semicontinuous φ satisfying $\varphi(1 + \varepsilon) < \infty$ and

$$\limsup_{n \rightarrow \infty} \frac{C_{\alpha, \mathcal{W}^{(n)}}}{t^{(n)}} = \varphi(\alpha) \quad \forall \alpha \in [1, 1 + \varepsilon] \quad (21)$$

then $\lim_{n \rightarrow \infty} P_e^{av(n)} = 1$ for any rate R sequence of codes for $R > \varphi(1)$.

Theorem 1 follows from Gallager's inner bound and Arimoto's outer bound. Theorem 1 establishes the equality of the channel capacity to the scaled order one Renyi capacity for a class of channels much larger than DSPCs and provides a sufficient condition for the existence of a strong converse. Theorem 1, however, does not claim either that the condition given in part (a) is necessary for the equality of the channel capacity to the scaled order one Renyi capacity, or that the condition given in part (b) is necessary for the existence of a strong converse.

The hypotheses of Theorem 1 are relatively easy to check if $\mathcal{W}^{(n)}$'s are product channels or product channels with feedback because of the additivity of the Renyi capacity established in Lemmas 1 and 2. In particular, let $\{\mathcal{W}_i\}_{i \in \mathbb{Z}^+}$ be a sequence of channels indexed by the positive integers such that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_{\alpha, \mathcal{W}_i} = \varphi(\alpha) \quad \forall \alpha \in [1 - \varepsilon, 1 + \varepsilon]. \quad (22)$$

If φ is a lower semicontinuous function on $[1 - \varepsilon, 1]$ then $C(\mathcal{W}_{\mathcal{T}_n}, n) = C(\mathcal{W}_{\vec{\mathcal{T}}_n}, n) = \varphi(1)$ for the index sets $\mathcal{T}_n = \{1, \dots, n\}$, as a result of Theorem 1-(a), Lemma 1 and Lemma 2. If φ is also an upper semicontinuous function on $[1, 1 + \varepsilon]$ that is finite at $(1 + \varepsilon)$ then we have a strong converse for both $\mathcal{W}_{\mathcal{T}_n}$ and $\mathcal{W}_{\vec{\mathcal{T}}_n}$.

²¹We do not need the limits given in (20) to exist for part (a). We only need $\limsup_{n \rightarrow \infty} C_{1, \mathcal{W}^{(n)}}/t^{(n)} = \varphi(1)$ and $\liminf_{n \rightarrow \infty} C_{\alpha, \mathcal{W}^{(n)}}/t^{(n)} = \varphi(\alpha)$ for all $\alpha \in (1 - \varepsilon, 1)$.

When the product channel is stationary, i.e. when $\mathcal{W}_i = \mathcal{W}$ for all i , equation (20) holds for $\varphi(\alpha) = C_{\alpha, \mathcal{W}}$ and continuity of φ on $(0, 1]$ follows from [35, Lemma 11-(c)]. Thus $C(\mathcal{W}_{\mathcal{T}_n}, n) = C(\mathcal{W}_{\vec{\mathcal{T}}_n}, n) = C_{1, \mathcal{W}}$. Furthermore, if $C_{\phi, \mathcal{W}} < \infty$ for an $\phi > 1$ then $C_{\alpha, \mathcal{W}}$ is continuous on $(0, \phi]$ by [35, Lemma 11-(f)] and the strong converse holds for both $\mathcal{W}_{\mathcal{T}_n}$ and $\mathcal{W}_{\vec{\mathcal{T}}_n}$.

Theorem 1 can be applied to certain sequences memoryless channels, with little effort, as well. Poisson channels described in [35, Examples 8 and 9] are not product channels, but they are memoryless, in the sense of Definition 5. For $T^{(n)} = t^{(n)} = n$ and finite b , Poisson channels in [35, Examples 8, 9, 10]²² satisfy the hypotheses for both parts of Theorem 1. The following expressions for the channel capacity of the sequences of channels follow from the corresponding expressions for the order one Renyi capacity of the individual channels given in [35, equations (104), (108), (115)].

$$\begin{aligned} C(\Lambda^{[n, a, b, \varrho]}, n) &= \frac{\varrho - a}{b - a} b \ln \frac{b}{\varrho} + \frac{b - \varrho}{b - a} a \ln \frac{a}{\varrho} \\ C(\Lambda^{[n, a, b, \leq \varrho]}, n) &= C(\Lambda^{[n, a, b, \varrho \wedge \varrho_{1, a, b}]}, n) \\ C(\Lambda^{[n, a, b, \geq \varrho]}, n) &= C(\Lambda^{[n, a, b, \varrho \vee \varrho_{1, a, b}]}, n) \\ C(\Lambda^{[n, a, b]}, n) &= C(\Lambda^{[n, a, b, \varrho_{1, a, b}]}, n) = \frac{b \frac{b-a}{b-a} a^{-\frac{b-a}{b-a}}}{e} - \frac{ab}{b-a} \ln \frac{b}{a} \end{aligned} \quad \text{where } \varrho_{1, a, b} = e^{-1} b^{\frac{b}{b-a}} a^{-\frac{a}{b-a}}$$

The closed form expressions for $C(\Lambda^{[n, a, b]}, n)$ and $C(\Lambda^{[n, a, b, \leq \varrho]}, n)$ were determined by Kabanov [33] and Davis [18], but only with weak converses. Strong converses for Poisson channels have been reported only for $a = 0$, i.e. zero dark current, cases: for $\Lambda^{[n, 0, b, \leq \varrho]}$ by Burnashev and Kutoyants [12], for $\Lambda^{[n, 0, b]}$ by Wagner and Anantharam [53]. Theorem 1 not only determines the closed form expressions for the capacity of the Poisson channels mentioned above, but also shows that strong converse holds for each one of them. Furthermore, the expressions for the Renyi capacities given in [35, equations (104), (108), (115)] and Arimoto's outer bound given in Lemma 5 imply bounds of the form $P_e^{av} \geq 1 - e^{-E_{sp}(R, \mathcal{W})}$ for R 's greater than $C_{1, \mathcal{W}}$ for any of the Poisson channels discussed in [35, Examples 8, 9, 10]. Such a bound was reported by Wagner and Anantharam [53] before, but only for $\Lambda^{[n, 0, b]}$ case.

Kabanov [33] proved that channel capacity of $\Lambda^{[n, a, b]}$ does not increase if it is extended by including Poisson processes whose intensity is dependent on the past arrivals. Lemma 2 provides us with a similar result, if we assume a non-zero delay. If there exists an $\epsilon > 0$ such that the intensity at any time τ depends only on the arrivals in $[0, \tau - \epsilon]$ for all Poisson processes in the extension then the Renyi capacity of any positive order is invariant under the extension by Lemma 2. Hence the channel capacity does not increase and the strong converse continues to hold. In order to recover Kabanov result which is valid for zero delay, i.e. $\epsilon = 0$, one needs to apply a martingale argument similar to the one used by Kabanov [33].

Hypotheses of Theorem 1 can be confirmed for certain channels with memory, as well. But formal statement and confirmation of those claims are beyond the scope of the current investigation. Before presenting the proof of Theorem 1, let us point out a sequence of channels violating both hypotheses of Theorem 1. Consider the channels described in [35, Example 3] for $t^{(n)} = \ln n$. Then,

$$\lim_{n \rightarrow \infty} \frac{C_{\alpha, \mathcal{W}^{(n)}}}{t^{(n)}} = \begin{cases} 0 & \alpha \in (0, 1) \\ 1 - \varrho & \alpha = 1 \\ 1 & \alpha \in (1, \infty) \end{cases}.$$

Theorem 1 is mute about the channel capacity of this sequence of channels. But $\lim_{n \rightarrow \infty} \frac{M^{(n)}}{L^{(n)}} = 1$ for any reliable sequence because $P_e^{av} \geq \frac{M-L}{M} \varrho$ for any (M, L) channel code on $\mathcal{W}^{(n)}$. Thus $C(\mathcal{W}^{(n)}, t^{(n)}) = 0$ for any $t^{(n)}$ such that $\lim_{n \rightarrow \infty} t^{(n)} > 0$. Previously, this sequence of channels is used by Verdú and Han [52] in order to motivate their investigation of a general expression for the channel capacity.

Proof of Theorem 1:

- (1-a) Let us first prove the direct part. For any $R < \varphi(1)$ there exists an α_0 such that $\varphi(\alpha) > R$ for all $\alpha \in (\alpha_0, 1)$ because φ is a lower semicontinuous function. Thus, there exists an $\alpha_1 \in (1 - \epsilon, 1)$ such that $\varphi(\alpha_1) > R$. Furthermore, there exists an n_0 such that $\frac{C_{\alpha_1, \mathcal{W}^{(n)}}}{t^{(n)}} > \frac{\varphi(\alpha_1) + R}{2}$ for all $n \geq n_0$ because $\liminf_{n \rightarrow \infty} \frac{C_{\alpha_1, \mathcal{W}^{(n)}}}{t^{(n)}} = \varphi(\alpha_1)$. Consequently, for each $n \geq n_0$ there exists a $p^{(n)} \in \mathcal{P}(\mathcal{W}^{(n)})$ such that $\frac{I_{\alpha_1}(p^{(n)}; \mathcal{W}^{(n)})}{t^{(n)}} > \frac{\varphi(\alpha_1) + R}{2}$ because $C_{\alpha_1, \mathcal{W}^{(n)}} = \sup_{p \in \mathcal{P}(\mathcal{W}^{(n)})} I_{\alpha_1}(p^{(n)}; \mathcal{W}^{(n)})$. Then by equation (13) of Lemma 4, for each $n > n_0$ there exists a $(e^{\lfloor t^{(n)} R \rfloor}, 1)$ channel code on $\mathcal{W}^{(n)}$ satisfying

$$P_e^{av(n)} \leq e^{\frac{\alpha_1 - 1}{\alpha_1} \frac{\varphi(\alpha_1) - R}{2} t^{(n)}}.$$

Thus for any $R < \varphi(1)$, we have a rate R reliable sequence. Consequently $C(\mathcal{W}^{(n)}, t^{(n)}) \geq \varphi(1)$.

The converse part is trivial for the case $\varphi(1) = \infty$. For $\varphi(1) < \infty$ case, we show that a rate R sequence of codes for an $R > \varphi(1)$ can not be reliable. For any $R > \varphi(1)$, there exists an n_0 such that $\frac{C_{1, \mathcal{W}^{(n)}}}{t^{(n)}} < \frac{R + 2\varphi(1)}{3}$ for all $n \geq n_0$

²²To be precise in [35, Examples 9], we have determined the Renyi capacity of $\Lambda^{[n, a, b, < \varrho]}$ and $\Lambda^{[n, a, b, > \varrho]}$ rather than the Renyi capacity of $\Lambda^{[n, a, b, \leq \varrho]}$ and $\Lambda^{[n, a, b, \geq \varrho]}$. However, $\Lambda^{[n, a, b, \leq \varrho]}$ is in the closure of $\Lambda^{[n, a, b, < \varrho]}$ and $\Lambda^{[n, a, b, \geq \varrho]}$ is in the closure of $\Lambda^{[n, a, b, > \varrho]}$ for the topology of setwise convergence. Thus they have the same Renyi capacity and Renyi center by [35, Lemma 21-(b)].

because $\limsup_{n \rightarrow \infty} \frac{C_{1, \mathcal{W}^{(n)}}}{t^{(n)}} = \varphi(1)$. On the other hand for any rate R sequence of codes there exists an n_1 such that $\frac{1}{t^{(n)}} \ln \frac{M^{(n)}}{L^{(n)}} > \frac{2R + \varphi(1)}{3}$ for all $n \geq n_1$ because $\liminf_{n \rightarrow \infty} \frac{1}{t^{(n)}} \ln \frac{M^{(n)}}{L^{(n)}} = R$. Then by equation (15) of Lemma 5 for $\alpha = 1$, i.e. by Fano's inequality, we have

$$(1 - P_e^{av(n)}) \ln \frac{M^{(n)}}{L^{(n)}} \leq C_{1, \mathcal{W}^{(n)}} + \ln 2 \quad \Rightarrow \quad P_e^{av(n)} \geq \frac{R - \varphi(1)}{2R + \varphi(1)} - \frac{3 \ln 2}{(2R + \varphi(1))t^{(n)}} \quad \forall n > (n_0 \vee n_1).$$

There does not exist a rate R reliable sequence of codes for $R > \varphi(1)$. Consequently $C(\mathcal{W}^{(n)}, t^{(n)}) \leq \varphi(1)$.

(1-b) For any $R > \varphi(1)$ there exists an α_0 such that $\varphi(\alpha) < R$ for all $\alpha \in (1, \alpha_0)$ because φ is an upper semicontinuous function by the hypothesis. Thus there exists an $\alpha_1 \in (1, 1 + \varepsilon)$ such that $\varphi(\alpha_1) < R$. Furthermore, there exists an n_0 such that $\frac{C_{\alpha_1, \mathcal{W}^{(n)}}}{t^{(n)}} < \frac{R + 2\varphi(\alpha_1)}{3}$ for all $n \geq n_0$ because $\limsup_{n \rightarrow \infty} \frac{C_{\alpha_1, \mathcal{W}^{(n)}}}{t^{(n)}} = \varphi(\alpha_1)$. On the other hand for any rate R sequence of codes there exists an n_1 such that $\frac{1}{t^{(n)}} \ln \frac{M^{(n)}}{L^{(n)}} > \frac{2R + \varphi(\alpha_1)}{3}$ for all $n \geq n_1$ because $\liminf_{n \rightarrow \infty} \frac{1}{t^{(n)}} \ln \frac{M^{(n)}}{L^{(n)}} = R$. Then by equation (15) of Lemma 5 for $\alpha = \alpha_1$, we have

$$(1 - P_e^{av(n)})^{\alpha_1} \left(\frac{L^{(n)}}{M^{(n)}} \right)^{1 - \alpha_1} \leq e^{(\alpha_1 - 1)C_{\alpha_1, \mathcal{W}^{(n)}}} \quad \Rightarrow \quad P_e^{av(n)} \geq 1 - e^{-\frac{\alpha_1 - 1}{\alpha_1} \frac{R - \varphi(\alpha_1)}{3} t^{(n)}} \quad \forall n > (n_0 \vee n_1).$$

Thus $\lim_{n \rightarrow \infty} P_e^{av(n)} = 1$ for any rate R sequence of codes for $R > \varphi(1)$. ■

Convergence of the scaled Renyi capacities to a continuous function φ on an arbitrarily small interval around $\alpha = 1$ implies the equality of the channel capacity and the limit of the scaled order one Renyi capacities. One expects the uniform convergence of the scaled Renyi capacities on compact subsets of $(0, 1)$ to imply the asymptotic optimality of Gallager's inner bound. We do not have such a result for arbitrary sequences of channels, yet. It is, however, possible to establish the optimality of Gallager's inner bound—in terms exponential decay rate of the error probability with the block length—for a class of product channels that includes all SPCs. We do the same for DSPCs with feedback and for various Poisson channels in Section IV and Appendix B.

III. THE SPHERE PACKING BOUND FOR PRODUCT CHANNELS

Assumption 1. $\{\mathcal{W}_t\}_{t \in \mathbb{Z}^+}$ is an ordered sequence of channels such that maximum $C_{\frac{1}{2}, \mathcal{W}_t}$ among the first n \mathcal{W}_t 's is $O(\ln n)$.

$$\exists n_0 \in \mathbb{Z}^+, K \in \mathbb{R}^+ \quad \text{such that} \quad \max_{t \leq n} C_{\frac{1}{2}, \mathcal{W}_t} \leq K \ln(n) \quad \forall n \geq n_0.$$

Theorem 2. Let $\{\mathcal{W}_t\}_{t \in \mathbb{Z}^+}$ be a sequence of channels satisfying Assumption 1, ε be a positive real number, and α_0, α_1 be orders satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes on the product channels $\{\mathcal{W}_{[1, n]}\}_{n \in \mathbb{Z}^+}$ satisfying

$$C_{\alpha_1, \mathcal{W}_{[1, n]}} \geq \ln \frac{M_n}{L_n} \geq C_{\alpha_0, \mathcal{W}_{[1, n]}} + \varepsilon (\ln n)^2 \quad \forall n \geq n_0 \quad (23)$$

there exists a $\tau \in \mathbb{R}^+$ and an $n_1 \geq n_0$ such that

$$P_e^{av(n)} \geq n^{-\tau} e^{-E_{sp}(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]})} \quad \forall n \geq n_1. \quad (24)$$

The main aim of this section is to prove Theorem 2; we do so following the program put forward by Augustin in [8], as we understand it. In Section III-A, we define the order α tilted probability measure $v_\alpha^{w, q}$ for probability measures w, q and prove the continuity of $v_\alpha^{w, q}$ as a function of the order α for the case when q is changing continuously with the order α . In Section III-B, we bound the small deviation probability for sums of independent random variables using Berry Essen theorem. In Section III-C, we define the averaged Renyi center $q_{\alpha, \mathcal{W}}^\varepsilon$ as the average of Renyi centers around α and analyze the resulting averaged sphere packing exponent. In Section III-D, we derive parametric outer bounds for codes on arbitrary product channels and on certain product channels with feedback. In Section III-E, we prove Theorem 2 using one of the bounds established in Section III-D.

Before starting the proof of Theorem 2 in earnest, we make a brief digression to discuss the implications of Theorem 2 and to compare it with the results of Augustin in [8] and [9]. Theorem 2 and Gallager's inner bound, i.e. Lemma 4, determine the optimal $P_e^{av(n)}$ up to a polynomial factor for all sequences of product channels satisfying Assumption 1 because Gallager's inner bound implies that²³

$$P_e^{av(n)} \leq e^{\frac{1 - \alpha_0}{\alpha_0}} e^{-E_{sp}(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]})} \quad \text{if } \exists \alpha_0 \in \left[\frac{1}{1 + L_n}, 1 \right] \text{ such that } \ln \frac{M_n}{L_n} \in [C_{\alpha_0, \mathcal{W}_{[1, n]}}, C_{1, \mathcal{W}_{[1, n]}} - 1]. \quad (25)$$

If the sequence of channels satisfying Assumption 1 have component channels with bounded order $1/2$ Renyi capacity, i.e. if there exists a $K \in \mathbb{R}^+$ such that $\sup_t C_{\frac{1}{2}, \mathcal{W}_t} \leq K$, then we can bound τ in Theorem 2 from above, as well. But, our bounding techniques are too crude to recover the right polynomial factor.

Augustin derives four sphere packing bounds discussed below. They all include Assumption 2 in their hypothesis.

²³If $R \in [C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}} - \delta]$ for a $\delta > 0$ then $E_{sp}(R, \mathcal{W}) \leq E_{sp}(R + \delta, \mathcal{W}) + \frac{1 - \alpha}{\alpha} \delta$ by equation (12).

Assumption 2. $\{\mathcal{W}_t\}_{t \in \mathbb{Z}^+}$ is an ordered sequence of channels such that²⁴

$$\lim_{\alpha \downarrow 0} \sup_{n \in \mathbb{Z}^+} g(\alpha, n) = 0 \quad \text{where} \quad g(\alpha, n) = \frac{1}{n} [C_{\alpha, \mathcal{W}_{[1,n]}} - \lim_{\phi \downarrow 0} C_{\phi, \mathcal{W}_{[1,n]}}]. \quad (26)$$

- [8, Theorem 4.7b] establishes $P_e^{\max(n)} \geq e^{-e^{\frac{\kappa}{2}} \sqrt{32n}} e^{-E_{sp}(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1,n]})}$ for n large enough for any sequence of channels satisfying Assumption 2 and $\sup_{t \in \mathbb{Z}^+} C_{1, \mathcal{W}_t} < \infty$. As an asymptotic result [8, Theorem 4.7b] proves a claim weaker than Theorem 2 for a hypothesis stronger than Assumption 1.
- [8, Theorem 4.8] establishes a sphere packing bound with a polynomial prefactor for channels satisfying Assumptions 2 and 3.

Assumption 3. $\{\mathcal{W}_t\}_{t \in \mathbb{Z}^+}$ is an ordered sequence of channels with a universal constant K and a sequence of probability measures $\{\nu_t\}_{t \in \mathbb{Z}^+}$ such that

$$\frac{1}{K} \leq \frac{dw}{d\nu_t} \leq K \quad w - \text{a.e.} \quad \forall w \in \mathcal{W}_t. \quad (27)$$

Assumption 3 implies Assumption 1, but Assumption 1 does not imply Assumption 3. Consider for example the sequence of Poisson channels $\Lambda^{[n,a,b]}$ indexed by $n \in \mathbb{Z}^+$ for $\Lambda^{[T,a,b]}$ described in [35, equation (103)]. It satisfies Assumption 1 but violates Assumption 3. Thus [8, Theorem 4.8] is weaker than Theorem 2 because it establishes the same claim for a stronger hypothesis.

- [8, Theorem 4.7a] and [9, Theorem 31.4] establish $P_e^{\max(n)} \geq e^{-O(\sqrt{n}) - E_{sp}(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1,n]})}$ for n large enough for channels satisfying Assumption 2 and some other hypothesis. They are not comparable with Theorem 2 because their hypotheses are not comparable with Assumption 1. For channels satisfying Assumptions 1 and 2, [9, Theorem 31.4] can be improved to have a polynomial prefactor if one invokes [9, Lemma 31.2] instead of [9, Lemma 31.1].

Assumption 2 has to be included in the hypothesis in order to obtain asymptotic bounds from [9, Lemma 31.1] or [9, Lemma 31.2]. Lemma 12 does not need such an assumption. Results similar to [8, Theorem 4.7a] and [9, Theorem 31.4] can be obtained from Lemma 12 without including Assumption 2 in the hypothesis.

A. Tilting with A Family of Measures

Definition 10. Let α be positive real number and w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ such that $D_\alpha(w \| q) < \infty$ then the order α tilted probability measure $v_\alpha^{w,q}$ is given by

$$\frac{dv_\alpha^{w,q}}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(w \| q)} \left(\frac{dw}{d\nu} \right)^\alpha \left(\frac{dq}{d\nu} \right)^{1-\alpha}. \quad (28)$$

Tilted probability measures arise naturally in the asymptotic trade off between the exponents of the false alarm and missed detection probabilities in the binary hypothesis testing problem with independent samples. We use them in the same vein with the help of bounds given in the following.

Lemma 6. Let w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ such that $D_{1/2}(w \| q) < \infty$ then

$$\mathbf{E}_{v_\alpha^{w,q}} [X_\alpha^{w,q | \kappa}]^{\frac{1}{\kappa}} \leq 3^{\frac{1}{\kappa}} \frac{((1-\alpha)D_\alpha(w \| q)) \vee \kappa}{\alpha(1-\alpha)} \quad \forall \kappa \in (0, \infty), \alpha \in (0, 1) \quad (29)$$

where $X_\alpha^{w,q}$ is defined using the Radon-Nikodym derivative of the component of w that is absolutely continuous in q as follows

$$X_\alpha^{w,q} \triangleq \ln \frac{dw_\sim}{dq} - \mathbf{E}_{v_\alpha^{w,q}} \left[\ln \frac{dw_\sim}{dq} \right].$$

Proof: We denote $v_\alpha^{w,q}$ by v_α and $X_\alpha^{w,q}$ by X_α in order to avoid notational clutter. Then for all $\gamma \geq 0$ and $\kappa \geq 0$ we have

$$\begin{aligned} \int_{X_\alpha > \gamma} (X_\alpha)^\kappa v_\alpha(dy) &= \int_{X_\alpha > \gamma} (X_\alpha)^\kappa \frac{dv_\alpha}{dw} w(dy) &&= \int_{X_\alpha > \gamma} (X_\alpha)^\kappa e^{(\alpha-1)X_\alpha + D_1(v_\alpha \| w)} w(dy) \\ &\leq w(X_\alpha > \gamma) e^{D_1(v_\alpha \| w)} \sup_{x > \gamma} e^{-(1-\alpha)x} x^\kappa &&\stackrel{(a)}{\leq} w(X_\alpha > \gamma) e^{D_1(v_\alpha \| w) - \frac{1-\alpha}{\alpha} D_\alpha(w \| q)} \gamma^\kappa \\ \int_{X_\alpha < -\gamma} (-X_\alpha)^\kappa v_\alpha(dy) &= \int_{X_\alpha < -\gamma} (-X_\alpha)^\kappa \frac{dv_\alpha}{dq} q(dy) &&= \int_{X_\alpha < -\gamma} (-X_\alpha)^\kappa e^{\alpha X_\alpha + D_1(v_\alpha \| q)} q(dy) \\ &\leq q(X_\alpha < -\gamma) e^{D_1(v_\alpha \| q)} \sup_{x > \gamma} e^{-\alpha x} x^\kappa &&\stackrel{(b)}{\leq} q(X_\alpha < -\gamma) e^{D_1(v_\alpha \| q) - D_\alpha(w \| q)} \gamma^\kappa \end{aligned}$$

For (a) and (b) we have assumed that $\gamma = \frac{((1-\alpha)D_\alpha(w \| q)) \vee \kappa}{\alpha(1-\alpha)}$ and used the following inequity

$$\sup_{x > \gamma} e^{-\beta x} x^\kappa = \begin{cases} \left(\frac{\kappa}{\beta} \right)^\kappa & \gamma \leq \frac{\kappa}{\beta} \\ e^{-\beta \gamma} \gamma^\kappa & \gamma > \frac{\kappa}{\beta} \end{cases} \quad \beta > 0, \kappa \geq 0, \gamma \geq 0.$$

²⁴This assumption is given as equation (7) in [8] and Condition 31.3a in [9]. In [9], g is defined without $\frac{1}{n}$ factor; we believe that is a typo.

Then equation (29) follows from the identity $(1 - \alpha)D_\alpha(w \| q) = \alpha D_1(v_\alpha \| w) + (1 - \alpha)D_1(v_\alpha \| q)$. \blacksquare

One can show —using standard results on the continuity of integrals, such as [11, Corollary 2.8.7]— that $D_1(v_\alpha^{w,q} \| w)$ and $D_1(v_\alpha^{w,q} \| q)$ are continuous functions of α on $(0, \phi)$ for every w and q satisfying $D_\phi(w \| q) < \infty$. In fact, such a proof will ensure the same continuity results even when q is replaced by a parametric family of probability measures q_α provided that there exists a $\exists \nu \in \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ satisfying the following three conditions

- (i) $w \prec \nu$ and $\{q_\alpha : \alpha \in (0, \phi)\} \prec \nu$
- (ii) $\frac{dq_\alpha}{d\nu}$ is a continuous function of α on $(0, \phi)$ for ν almost every y .
- (iii) \exists a ν -integrable real valued function g satisfying $(\frac{dw}{d\nu})^\alpha (\frac{dq_\alpha}{d\nu})^{1-\alpha} \left(1 + \left|\ln \frac{dw}{d\nu} - \ln \frac{dq_\alpha}{d\nu}\right|\right) \leq g$ for all $\alpha \in (0, \phi)$.

If we restrict our attention to the orders between zero and one continuity of $D_1(v_\alpha^{w,q_\alpha} \| w)$ and $D_1(v_\alpha^{w,q_\alpha} \| q_\alpha)$ follows from a weaker hypothesis: continuity of q_α in the total variation topology.²⁵

Lemma 7. *Let $(\mathcal{Y}, \mathcal{Y})$ be a measurable space, $q_\alpha : (0, 1) \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ be a continuous function for the total variation topology on $\mathcal{P}(\mathcal{Y}, \mathcal{Y})$ and w be a probability measure on $(\mathcal{Y}, \mathcal{Y})$ such that $D_\alpha(w \| q_\alpha) < \infty$ for all $\alpha \in (0, 1)$. Then*

- (a) $v_\alpha^{w,q_\alpha} : (0, 1) \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ is a continuous function for the total variation topology on $\mathcal{P}(\mathcal{Y}, \mathcal{Y})$.
- (b) $D_\alpha(w \| q_\alpha)$, $D_1(v_\alpha^{w,q_\alpha} \| w)$ and $D_1(v_\alpha^{w,q_\alpha} \| q_\alpha)$ are continuous functions of α from $(0, 1)$ to $[0, \infty)$.

If $C_{1/2, \mathcal{W}} < \infty$ then $q_{\alpha, \mathcal{W}}$ satisfies the hypothesis of Lemma 7 for all $w \in \mathcal{W}$ because $C_{\alpha, \mathcal{W}} < \infty$ for all $\alpha \in (0, 1)$ by [35, Lemma 11-(e)], $D_\alpha(w \| q_{\alpha, \mathcal{W}}) \leq C_{\alpha, \mathcal{W}}$ by [35, Theorem 1] and $q_{\alpha, \mathcal{W}}$ is continuous in α for the total variation topology on $\mathcal{P}(\mathcal{Y}, \mathcal{Y})$ by [35, Lemma 17]. We believe $q_{\alpha, \mathcal{W}}$ satisfies the monotonicity property described in [35, Conecture 1]. If that is the case, we can establish the continuity of $v_\alpha^{w, q_{\alpha, \mathcal{W}}}$, $D_\alpha(w \| q_{\alpha, \mathcal{W}})$, $D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| w)$ and $D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| q_{\alpha, \mathcal{W}})$ using standard results on the continuity of integrals, such as [11, Corollary 2.8.7].

Proof of Lemma 7: Any function g on $(0, 1)$ is continuous²⁶ if $\lim_{i \rightarrow \infty} g(\alpha^{(i)}) = g(\lim_{i \rightarrow \infty} \alpha^{(i)})$ for every convergent sequence $\{\alpha^{(i)}\}_{i \in \mathbb{Z}^+}$ by [34, Theorem 21.3] because $(0, 1)$ is metrizable. Let $\{\alpha^{(i)}\}_{i \in \mathbb{Z}^+}$ be a convergent sequence such that $\lim_{i \rightarrow \infty} \alpha^{(i)} = \alpha$ and ν be

$$\nu = \frac{w}{4} + \frac{q_\alpha}{4} + \frac{1}{2} \sum_{i \in \mathbb{Z}^+} \frac{q_{\alpha^{(i)}}}{2^i}.$$

Instead of working with measures as members of $\mathcal{M}(\mathcal{Y}, \mathcal{Y})$ for the total variation topology, we work with corresponding Radon-Nikodym derivatives with respect to ν as members of $\mathcal{L}_1(\nu)$. We can do so because all of the measures we are considering are absolutely continuous with respect to ν and for any sequence $\{f^{(i)}\}_{i \in \mathbb{Z}^+} \subset \mathcal{L}_1(\nu)$, $\{f^{(i)}\} \xrightarrow{\mathcal{L}_1(\nu)} f$ iff corresponding sequence of measures $\{f^{(i)}\nu\}_{i \in \mathbb{Z}^+}$ converges to $f\nu$ in $\mathcal{M}(\mathcal{Y}, \mathcal{Y})$ for the total variation topology.

For any finite signed measure μ such that $\mu \prec \nu$, we denote its Radon-Nikodym derivative with respect to ν by f_μ :

$$f_\mu = \frac{d\mu}{d\nu} \quad \forall \mu \in \mathcal{M}(\mathcal{Y}, \mathcal{Y}) \text{ such that } \mu \prec \nu. \quad (30)$$

In order to avoid notational clutter, we make an exception and denote the Radon-Nikodym derivative of v_α^{w, q_α} by f_{v_α} rather than $f_{v_\alpha^{w, q_\alpha}}$.

(7-a) For all $\alpha \in (0, 1)$ let f_{s_α} be $f_{s_\alpha} \triangleq f_w^\alpha f_{q_\alpha}^{1-\alpha}$. Using triangle equality we get:

$$\left| f_{s_\alpha} - f_{s_{\alpha^{(i)}}} \right| \leq \left| f_{s_\alpha} - f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}} \right| + \left| f_w^{\alpha^{(i)}} \left| f_{q_\alpha}^{1-\alpha^{(i)}} - f_{q_{\alpha^{(i)}}}^{1-\alpha^{(i)}} \right| \right|. \quad (31)$$

Note that

- $\{f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}}\}_{i \in \mathbb{Z}^+}$ is uniformly integrable because $\int_{\mathcal{E}} f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}} \nu(dy) \leq w(\mathcal{E})^{\alpha^{(i)}} q_\alpha(\mathcal{E})^{1-\alpha^{(i)}}$ by Holder's inequality and $w(\mathcal{E})^{\alpha^{(i)}} q_\alpha(\mathcal{E})^{1-\alpha^{(i)}} \leq w(\mathcal{E}) + q_\alpha(\mathcal{E})$.
- $\{f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}}\}_{i \in \mathbb{Z}^+} \xrightarrow{\nu} f_{s_\alpha}$ because almost everywhere convergence implies convergence in measure for finite measures by [11, Theorem 2.2.3] and $\{f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}}\}_{i \in \mathbb{Z}^+} \xrightarrow{\nu-a.e.} f_{s_\alpha}$ by definition.

Then by Lebesgue-Vitali convergence theorem [11, 4.5.4] we have

$$\lim_{i \rightarrow \infty} \int \left| f_{s_\alpha} - f_w^{\alpha^{(i)}} f_{q_\alpha}^{1-\alpha^{(i)}} \right| \nu(dy) = 0. \quad (32)$$

Using the derivative test one can confirm that $[(x + \tau)^\beta - x^\beta]$ is a decreasing function of x for $x \geq 0$, $\tau \geq 0$, $\beta \in (0, 1)$. Thus $(x + \tau)^\beta - x^\beta \leq \tau^\beta$ for any $x \geq 0$, $\tau \geq 0$, $\beta \in (0, 1)$. Then using Holder's inequality we get,

$$\begin{aligned} \int f_w^{\alpha^{(i)}} \left| f_{q_\alpha}^{1-\alpha^{(i)}} - f_{q_{\alpha^{(i)}}}^{1-\alpha^{(i)}} \right| \nu(dy) &\leq \int f_w^{\alpha^{(i)}} \left| f_{q_\alpha} - f_{q_{\alpha^{(i)}}} \right|^{1-\alpha^{(i)}} \nu(dy) \\ &\leq \left(\int f_w \nu(dy) \right)^{\alpha^{(i)}} \left(\int \left| f_{q_\alpha} - f_{q_{\alpha^{(i)}}} \right| \nu(dy) \right)^{1-\alpha^{(i)}} \end{aligned}$$

²⁵Continuity in total variation topology does not imply the continuity of the corresponding Radon-Nikodym derivatives, see [35, footnote 32 on page 24].

²⁶Convergence implied by the equality $\lim_{i \rightarrow \infty} g(\alpha^{(i)}) = g(\lim_{i \rightarrow \infty} \alpha^{(i)})$ is determined by the topology on the range space of the function g .

Furthermore $\left\{ \int \left| f_{q_\alpha} - f_{q_{\alpha^{(i)}}} \right| \nu(dy) \right\} \rightarrow 0$ because $\{f_{q_{\alpha^{(i)}}}\} \xrightarrow{\mathcal{L}_1(\nu)} f_{q_\alpha}$. Then using $\{\alpha^{(i)}\} \rightarrow \alpha$ and $\alpha \in (0, 1)$ we get

$$\lim_{i \rightarrow \infty} \int f_w^{\alpha^{(i)}} \left| f_{q_\alpha}^{1-\alpha^{(i)}} - f_{q_{\alpha^{(i)}}}^{1-\alpha^{(i)}} \right| \nu(dy) = 0. \quad (33)$$

Then $\{f_{s_{\alpha^{(i)}}}\} \xrightarrow{\mathcal{L}_1(\nu)} f_{s_\alpha}$ follows from equations (31), (32) and (33). Thus $s_\alpha : (0, 1) \rightarrow \mathcal{M}^+(\mathcal{Y}, \mathcal{Y})$ is continuous in α for the total variation topology on $\mathcal{M}^+(\mathcal{Y}, \mathcal{Y})$. Then $\|s_\alpha\| = \int f_{s_\alpha} \nu(dy)$ is continuous in α , as well. Furthermore, $\|s_\alpha\| > 0$ because $\|s_\alpha\| = e^{(\alpha-1)D_\alpha(w\|q_\alpha)}$ and $D_\alpha(w\|q) < \infty$ by the hypothesis of the lemma. On the other hand, $f_{v_\alpha} = f_{s_\alpha}/\|s_\alpha\|$. Then using the triangle inequality, we get

$$\left| f_{v_\alpha} - f_{v_{\alpha^{(i)}}} \right| \leq \frac{1}{\|s_\alpha\|} \left| f_{s_\alpha} - f_{s_{\alpha^{(i)}}} \right| + \left| \frac{\|s_\alpha\| - \|s_{\alpha^{(i)}}\|}{\|s_\alpha\|} \right|.$$

Since $\|s_\alpha\| > 0$, continuity of $\|s_\alpha\|$ in α and $\{f_{s_{\alpha^{(i)}}}\} \xrightarrow{\mathcal{L}_1(\nu)} f_{s_\alpha}$ implies that

$$\lim_{i \rightarrow \infty} \int \left| f_{v_\alpha}^{1-\alpha^{(i)}} - f_{v_{\alpha^{(i)}}}^{1-\alpha^{(i)}} \right| \nu(dy) = 0.$$

Then $\{f_{v_{\alpha^{(i)}}}\} \xrightarrow{\mathcal{L}_1(\nu)} f_{v_\alpha}$ and $v_\alpha^{w, q_\alpha} : (0, 1) \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ is continuous in α for the total variation topology.

(7-b) Note that $\|s_\alpha\| > 0$ by the hypothesis because $D_\alpha(w\|q_\alpha) = \frac{1}{\alpha-1} \ln \|s_\alpha\|$. Furthermore, $D_\alpha(w\|q_\alpha)$ is continuous in α because product and composition of continuous functions are continuous.

$D_1(v_\alpha^{w, q_\alpha} \| w)$ and $D_1(v_\alpha^{w, q_\alpha} \| q_\alpha)$ are both lower semicontinuous in α because Renyi divergence is jointly lower semicontinuous in its arguments for the topology of setwise convergence by [35, Lemma 7] and v_α^{w, q_α} and q_α are continuous in the topology of setwise convergence.

$D_1(v_\alpha^{w, q_\alpha} \| w)$ is upper semicontinuous in α because $D_\alpha(w\|q_\alpha)$ is continuous in α , $D_1(v_\alpha^{w, q_\alpha} \| q_\alpha)$ is lower semicontinuous in α and $D_1(v_\alpha^{w, q_\alpha} \| w) = \frac{1-\alpha}{\alpha} D_\alpha(w\|q_\alpha) - \frac{1-\alpha}{\alpha} D_1(v_\alpha^{w, q_\alpha} \| q_\alpha)$.

Then $D_1(v_\alpha^{w, q_\alpha} \| w)$ is continuous in α because it is both lower semicontinuous and upper semicontinuous in α .

Expressing $D_1(v_\alpha^{w, q_\alpha} \| q_\alpha)$ in terms of $D_1(v_\alpha^{w, q_\alpha} \| w)$ and following a similar reasoning, we deduce that $D_1(v_\alpha^{w, q_\alpha} \| q_\alpha)$ is continuous in α , as well. ■

B. A Lower Bound on the Probability of Small Deviations via Berry Essen Theorem

In this section we bound from below the probability of having a small deviation from the mean for sums of independent random variables using Berry Essen theorem. Let us start with recalling Berry Essen theorem.

Lemma 8 (Berry Essen Theorem [10],[22],[49]). *Let $\{X_t\}_{t \in \mathbb{Z}^+}$ be independent random variables such that*

$$\mathbf{E}[X_t] = 0 \quad \forall t \quad \text{and} \quad g_2 \in (0, \infty) \quad \text{where} \quad g_\kappa = \left(\sum_{t=1}^n \mathbf{E}[|X_t|^\kappa] \right)^{\frac{1}{\kappa}}.$$

Then there exists an absolute constant $\omega \leq 0.5600$ such that

$$\left| \mathbf{P} \left[\sum_{t=1}^n X_t < \tau g_2 \right] - \Phi(\tau) \right| \leq \omega \left(\frac{g_3}{g_2} \right)^3 \quad \text{where} \quad \Phi(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-\frac{x^2}{2}} \lambda(dx).$$

Lemma 9. *Let $\{X_t\}_{t \in \mathbb{Z}^+}$ be independent zero mean random variables then*

$$\mathbf{P} \left[\left| \sum_{t=1}^n X_t \right| < 3g_\kappa \right] \geq \frac{1}{2\sqrt{n}} \quad \forall n \in \mathbb{Z}^+, \kappa \geq 3. \quad (34)$$

Augustin derived a similar bound in [9]; proof of Lemma 9 is similar to the proof of that bound, i.e. [9, Theorem 18.2].

Proof of Lemma 9: If $g_2\sqrt{2} \leq 3g_\kappa$ then

$$\begin{aligned} \mathbf{P} \left[\left| \sum_{t=1}^n X_t \right| \leq 3g_\kappa \right] &\geq 1 - \mathbf{P} \left[\left| \sum_{t=1}^n X_t \right| > g_2\sqrt{2} \right] \\ &\geq \frac{1}{2}. \end{aligned}$$

Thus inequality (34) holds. Hence we assume that $g_2\sqrt{2} > 3g_\kappa$ for the rest of the proof. By Berry Essen theorem,

$$\mathbf{P} \left[\left| \sum_{t=1}^n X_t \right| \leq 3g_\kappa \right] \geq 2 \left[\int_0^{\frac{3g_\kappa}{g_2}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \lambda(dx) - \omega \left(\frac{g_3}{g_2} \right)^3 \right]. \quad (35)$$

By Holder's inequality we have

$$\mathbf{E} \left[\sum_{t=1}^n |X_t|^3 \right] \leq \mathbf{E} \left[\sum_{t=1}^n |X_t|^\kappa \right]^{\frac{1}{\kappa-2}} \mathbf{E} \left[\sum_{t=1}^n |X_t|^2 \right]^{\frac{\kappa-3}{\kappa-2}}.$$

Thus

$$\left(\frac{g_3}{g_2}\right)^3 \leq \left(\frac{g_\kappa}{g_2}\right)^{\frac{\kappa}{\kappa-2}}. \quad (36)$$

Since $g_2\sqrt{2} > 3g_\kappa$ we also have

$$\left(\frac{g_\kappa}{g_2}\right)^{\frac{\kappa}{\kappa-2}} \leq \left(\frac{g_\kappa}{g_2}\right). \quad (37)$$

On the other hand, as a result of the convexity of the exponential function and Jensen's inequality we have

$$\begin{aligned} \int_0^\xi \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \lambda(dx) &\geq \frac{\xi}{\sqrt{2\pi}} e^{-\frac{1}{\xi} \int_0^\xi \frac{x^2}{2} \lambda(dx)} \\ &\geq \frac{e^{-\frac{1}{3}}}{\sqrt{2\pi}} \xi \end{aligned} \quad \forall \xi \in [0, \sqrt{2}]. \quad (38)$$

Using equations (35), (36), (37) and (38) we get,

$$\mathbf{P}\left[\left|\sum_{t=1}^n \mathbf{X}_t\right| \leq 3g_\kappa\right] \geq 2 \left[\frac{e^{-\frac{1}{3}}}{\sqrt{2\pi}} 3 - 0.56\right] \frac{g_\kappa}{g_2} \geq 0.595 \frac{g_\kappa}{g_2}. \quad (39)$$

In order to bound $\frac{g_\kappa}{g_2}$ we use the Jensen's inequality and the concavity of the functions x^α for $\alpha \in (0, 1]$.

$$\mathbf{E}\left[\sum_{t=1}^n \frac{1}{n} |\mathbf{X}_t|^\ell\right]^{\frac{1}{\ell}} = \mathbf{E}\left[\sum_{t=1}^n \frac{1}{n} |\mathbf{X}_t|^{\kappa \frac{\ell}{\kappa}}\right]^{\frac{1}{\ell}} \leq \mathbf{E}\left[\sum_{t=1}^n \frac{1}{n} |\mathbf{X}_t|^\kappa\right]^{\frac{1}{\kappa}} \quad \forall \ell \leq \kappa$$

Then,

$$\frac{g_\kappa}{g_\ell} \geq n^{\frac{\ell-\kappa}{\ell\kappa}} \geq n^{-\frac{1}{\ell}} \quad \forall \ell \leq \kappa. \quad (40)$$

Equation (34) follows from (39) and (40). ■

C. Averaged Renyi Center, Averaged Renyi Capacity and Averaged Sphere Packing Exponent

For any $\epsilon \in (0, 1)$, we define the averaged Renyi center $q_{\alpha, \mathcal{W}}^\epsilon$ as a Lipschitz continuous²⁷ function of the order α with Lipschitz constant $1/\epsilon$. Then we define the averaged Renyi capacity $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ and the averaged sphere packing exponent $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ and show —using the convexity and monotonicity properties of the Renyi divergence— that $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ and $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ differ from the corresponding quantities $C_{\alpha, \mathcal{W}}$ and $E_{sp}(R, \mathcal{W})$ at most by a factor proportional to ϵ . In the following subsection, we derive an outer bound for codes on product channels in terms of $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ using $q_{\alpha, \mathcal{W}}^\epsilon$ and $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$.

If $C_{x, \mathcal{W}}$ is finite for a positive x then $C_{\alpha, \mathcal{W}}$ is finite for all orders α in $(0, 1)$ by [35, Lemma 11]. On the other hand, if $C_{\alpha, \mathcal{W}}$ is finite then there exists a unique order α Renyi center $q_{\alpha, \mathcal{W}}$, by [35, Theorem 1]. Furthermore, $q_{\cdot, \mathcal{W}} : (0, 1) \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ is a continuous function for the total variation topology on $\mathcal{P}(\mathcal{Y}, \mathcal{Y})$, by [35, Lemma 17]. As a result, $q_{\cdot, \mathcal{W}}(\mathcal{E}) : (0, 1) \rightarrow [0, 1]$ is a continuous²⁸ and hence a $(\mathcal{B}((0, 1)), \mathcal{B}([0, 1]))$ -measurable function for any $\mathcal{E} \in \mathcal{Y}$. On the other hand, Renyi center $q_{\alpha, \mathcal{W}}$ is a probability measure on $(\mathcal{Y}, \mathcal{Y})$ for all α by definition. Thus $q_{\cdot, \mathcal{W}}$ is a transition probability from $((0, 1), \mathcal{B}((0, 1)))$ to $(\mathcal{Y}, \mathcal{Y})$. We define the averaged Renyi center using this transition probability.

Definition 11. Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a channel such that $C_{1/2, \mathcal{W}} < \infty$ and α and ϵ be real numbers in $(0, 1)$. Then the averaged Renyi center $q_{\alpha, \mathcal{W}}^\epsilon$ is the \mathcal{Y} marginal of the probability measure $u_{\alpha, \epsilon} \circ q_{\cdot, \mathcal{W}}$ where $u_{\alpha, \epsilon}$ is the uniform probability distribution on $(\alpha - \epsilon\alpha, \alpha + \epsilon(1 - \alpha))$:

$$q_{\alpha, \mathcal{W}}^\epsilon \triangleq \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha + \epsilon(1 - \alpha)} q_{z, \mathcal{W}} \lambda(dz) \quad (41)$$

where λ is the Lebesgue measure.

For channels with certain symmetries such as $\mathcal{W}^{(2)}$ of [35, Example 1] or channels given in [35, Examples 4-7], $q_{\alpha, \mathcal{W}}$ is the same probability measure for all α for which it exists. For certain other channels, such as $\mathcal{W}^{(1)}$ of [35, Example 1], $q_{\alpha, \mathcal{W}}$ is same for all α on an interval. In these cases $\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) = C_{\alpha, \mathcal{W}}$ because $q_{\alpha, \mathcal{W}}^\epsilon = q_{\alpha, \mathcal{W}}$. However, we can not assert the equality of $q_{\alpha, \mathcal{W}}$ and $q_{\alpha, \mathcal{W}}^\epsilon$ in general and whenever $q_{\alpha, \mathcal{W}}^\epsilon \neq q_{\alpha, \mathcal{W}}$ the supremum $\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon)$ is strictly greater than $C_{\alpha, \mathcal{W}}$. In particular, by Erven Harremoës bound [35, Lemma 16] we have

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \geq C_{\alpha, \mathcal{W}} + D_\alpha(q_{\alpha, \mathcal{W}} \| q_{\alpha, \mathcal{W}}^\epsilon).$$

²⁷We use the usual topology of the real numbers on the domain and the total variation topology on $\mathcal{P}(\mathcal{Y}, \mathcal{Y})$.

²⁸Continuity of the Renyi center for the topology of setwise convergence is sufficient for ensuring the continuity of $q_{\alpha, \mathcal{W}}(\mathcal{E})$ as a function of α , for all $\mathcal{E} \in \mathcal{Y}$. We do not need the continuity of the Renyi center for the total variation topology.

Lemma 10 bounds $\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon)$ from above in terms of an integral of the Renyi capacity, which converges to $C_{\alpha, \mathcal{W}}$ as ϵ converges to zero for any $\alpha \in (0, 1)$.

Lemma 10. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a channel such that $C_{1/2, \mathcal{W}} < \infty$ and α and ϵ be real numbers in $(0, 1)$. Then*

$$\sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon \leq \frac{C_{1/2, \mathcal{W}}}{(1-\alpha)(1-\epsilon)} \quad (42)$$

where $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ is called the averaged Renyi capacity and given by

$$\tilde{C}_{\alpha, \mathcal{W}}^\epsilon \triangleq \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left[1 \vee \left(\frac{\alpha}{1-\alpha} \frac{1-z}{z} \right) \right] C_{z, \mathcal{W}} \lambda(dz). \quad (43)$$

Since $C_{\alpha, \mathcal{W}}$ is a continuous function of α on $(0, 1)$ for any channel \mathcal{W} by [35, Lemma 11-(c)], we have

$$\lim_{\epsilon \downarrow 0} \tilde{C}_{\alpha, \mathcal{W}}^\epsilon = C_{\alpha, \mathcal{W}} \quad \forall \alpha \in (0, 1).$$

Furthermore, the additivity of $C_{\alpha, \mathcal{W}}$ for product channels, i.e. Lemma 1, implies the additivity of $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ for product channels:

$$\tilde{C}_{\alpha, \mathcal{W}_{\mathcal{T}}}^\epsilon = \sum_{t \in \mathcal{T}} \tilde{C}_{\alpha, \mathcal{W}_t}^\epsilon. \quad (44)$$

Lemma 1 also states that $q_{\alpha, \mathcal{W}_{\mathcal{T}}} = \prod_{t \in \mathcal{T}} q_{\alpha, \mathcal{W}_t}$. The averaged Renyi center $q_{\alpha, \mathcal{W}_{\mathcal{T}}}^\epsilon$, however, does not satisfy such a product structure, in general.

Proof of Lemma 10: As a result of the convexity of the Renyi divergence in its second argument, i.e. [35, Lemma 9-(d)], and Jensen's inequality we have

$$D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \frac{D_\alpha(w \| q_{z, \mathcal{W}})}{\epsilon} \lambda(dz). \quad (45)$$

For any probability measure w , $D_\alpha(w \| q_{z, \mathcal{W}})$ is an increasing function of the order α , by [35, Lemma 9-(a)]. Furthermore, $D_\alpha(w \| q_{z, \mathcal{W}}) = \frac{\alpha}{1-\alpha} D_{1-\alpha}(q_{z, \mathcal{W}} \| w)$ for $\alpha \in (0, 1)$, by the definition of the Renyi divergence. Thus

$$\begin{aligned} D_\alpha(w \| q_{z, \mathcal{W}}) &\leq \mathbb{1}_{\{z \geq \alpha\}} D_z(w \| q_{z, \mathcal{W}}) + \mathbb{1}_{\{z < \alpha\}} \frac{\alpha}{1-\alpha} \frac{1-z}{z} D_z(w \| q_{z, \mathcal{W}}) \\ &= \left(1 \vee \frac{\alpha}{1-\alpha} \frac{1-z}{z} \right) D_z(w \| q_{z, \mathcal{W}}) \quad \forall w \in \mathcal{W}. \end{aligned} \quad (46)$$

Recall that $D_z(w \| q_{z, \mathcal{W}}) \leq C_{z, \mathcal{W}}$ for all $w \in \mathcal{W}$ by [35, Theorem 1]. Then using equations (45), (46) and the definition of $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ given in equation (43) we get

$$D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon \quad \forall w \in \mathcal{W}.$$

Then $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ is also an upper bound on the supremum of $D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon)$ for $w \in \mathcal{W}$.

In order to bound $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$, recall that $C_{\alpha, \mathcal{W}}$ is increasing in α by [35, Lemma 11-(a)] and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is decreasing in α on $(0, 1)$ by [35, Lemma 11-(c)]. Thus we have,

$$C_{z, \mathcal{W}} \leq \frac{z}{1-z} C_{1/2, \mathcal{W}} \mathbb{1}_{\{z > 1/2\}} + C_{1/2, \mathcal{W}} \mathbb{1}_{\{z \leq 1/2\}} \quad \forall z \in (0, 1).$$

Then by the definition of $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ given in (43) we have

$$\begin{aligned} \tilde{C}_{\alpha, \mathcal{W}}^\epsilon &\leq \frac{C_{1/2, \mathcal{W}}}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left(1 \vee \frac{\alpha}{1-\alpha} \frac{1-z}{z} \right) \left(1 \vee \frac{z}{1-z} \right) \lambda(dz) \leq \frac{C_{1/2, \mathcal{W}}}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \left(\frac{1}{1-z} \vee \frac{\alpha}{1-\alpha} \frac{1}{z} \right) \lambda(dz) \\ &\leq \frac{C_{1/2, \mathcal{W}}}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \frac{1}{(1-\alpha)(1-\epsilon)} \lambda(dz). \end{aligned}$$

Using the monotonicity of $C_{\alpha, \mathcal{W}}$, i.e. [35, Lemma 11-(a)], one can show that $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon \geq (\alpha \wedge \frac{2\alpha-1}{2\epsilon}) \frac{C_{1/2, \mathcal{W}}}{2(1-\alpha)}$. Thus for any channel \mathcal{W} with a finite but non-zero $C_{1/2, \mathcal{W}}$, we have²⁹ ■

$$\lim_{\alpha \uparrow 1} \tilde{C}_{\alpha, \mathcal{W}}^\epsilon = \infty \quad \forall \epsilon \in (0, 1).$$

Definition 12. Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a channel satisfying $C_{1/2, \mathcal{W}} \in (0, \infty)$, $R \in (0, \infty)$, and $\epsilon \in (0, 1)$. Then the averaged sphere packing exponent $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is given by

$$\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \triangleq \sup_{\alpha \in (0, 1)} \frac{1-\alpha}{\alpha} \left(\tilde{C}_{\alpha, \mathcal{W}}^\epsilon - R \right). \quad (47)$$

²⁹This is true even when $C_{1, \mathcal{W}}$ is finite. Thus, $C_{\alpha, \mathcal{W}}$ can be approximated by $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ uniformly only on compact subsets of $(0, 1)$, but not on $(0, 1)$.

$\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is decreasing and convex in R on $(0, \infty)$ because it is the pointwise supremum of decreasing and convex functions of R . One can show that $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ is increasing and continuous in α on $(0, 1)$ for any $\epsilon \in (0, 1)$ using the continuity and monotonicity of $C_{\alpha, \mathcal{W}}$ in α on $(0, 1)$. Since we do not need this observation in our analysis, we leave its proof to the interested reader. Using the monotonicity of $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ one can also show that $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is finite and continuous in R on $(\lim_{\alpha \downarrow 0} \tilde{C}_{\alpha, \mathcal{W}}^\epsilon, \infty)$.

Lemma 11. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a channel satisfying $C_{1/2, \mathcal{W}} \in (0, \infty)$, $\phi \in (0, 1)$, $R \in [C_{\phi, \mathcal{W}}, C_{1, \mathcal{W}}]$ and $\epsilon \in (0, \phi)$. Then*

$$0 \leq \tilde{E}_{sp}^\epsilon(R, \mathcal{W}) - E_{sp}(R, \mathcal{W}) \leq \frac{\epsilon}{\phi - \epsilon} (R \vee E_{sp}(R, \mathcal{W})) \leq \frac{\epsilon}{\phi - \epsilon} \frac{R}{\phi}. \quad (48)$$

If $C_{1, \mathcal{W}} < \infty$ then for any $R \in [C_{1, \mathcal{W}}, \infty)$ and $\epsilon \in (0, \phi)$ we have

$$\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \leq \frac{\epsilon}{1 - \epsilon} C_{1, \mathcal{W}}. \quad (49)$$

Proof of Lemma 11: $C_{\alpha, \mathcal{W}} \leq \sup_{w \in \mathcal{W}} D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ by Lemma 10 and [35, Theorem 1]. Then as a result of the expressions for $E_{sp}(R, \mathcal{W})$ given in equation (12) and the definition of $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ we have

$$E_{sp}(R, \mathcal{W}) \leq \tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \quad \forall R \in (0, C_{1, \mathcal{W}}). \quad (50)$$

Let us proceed with bounding $\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) - E_{sp}(R, \mathcal{W})$ from above for $R \in [C_{\phi, \mathcal{W}}, C_{1, \mathcal{W}}]$.

$$\begin{aligned} \frac{1-\alpha}{\alpha} (\tilde{C}_{\alpha, \mathcal{W}}^\epsilon - R) &= \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha + \epsilon(1-\alpha)} \left(\frac{1-\alpha}{\alpha} \vee \frac{1-z}{z} \right) C_{z, \mathcal{W}} \lambda(dz) - \frac{1-\alpha}{\alpha} R \\ &= \frac{1}{\epsilon} \frac{1-\alpha}{\alpha} \int_{\alpha}^{\alpha + \epsilon(1-\alpha)} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha} \frac{1-z}{z} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{R}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha} \frac{\alpha - z}{z\alpha} \lambda(dz) \\ &\leq \frac{1}{\epsilon} \frac{1-\alpha}{\alpha} \int_{\alpha}^{\alpha + \epsilon(1-\alpha)} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha} \frac{1-z}{z} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{\epsilon}{1-\epsilon} R \end{aligned} \quad (51)$$

We bound $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ by bounding the expression in (51) separately on two intervals for α .

Renyi capacity is an increasing function of the order by [35, Lemma 11-(a)], then for any $R \geq C_{\phi, \mathcal{W}}$ we have

$$\frac{1-\alpha}{\alpha} (\tilde{C}_{\alpha, \mathcal{W}}^\epsilon - R) \leq \frac{\epsilon}{1-\epsilon} R \quad \forall \alpha \in (0, \frac{\phi - \epsilon}{1 - \epsilon}]. \quad (52)$$

In order to bound the expression in (51) for $\alpha \in [\frac{\phi - \epsilon}{1 - \epsilon}, 1)$, we use the fact that $\frac{1-z}{z} (C_{z, \mathcal{W}} - R) \leq E_{sp}(R, \mathcal{W})$ for all $z \in (0, 1)$.

$$\begin{aligned} \frac{1-\alpha}{\alpha} (\tilde{C}_{\alpha, \mathcal{W}}^\epsilon - R) &\leq \frac{1}{\epsilon} \frac{1-\alpha}{\alpha} \int_{\alpha}^{\alpha + \epsilon(1-\alpha)} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha} \frac{1-z}{z} (C_{z, \mathcal{W}} - R) \lambda(dz) + \frac{\epsilon}{1-\epsilon} R \\ &\leq \frac{1}{\epsilon} \frac{1-\alpha}{\alpha} \int_{\alpha}^{\alpha + \epsilon(1-\alpha)} \frac{z}{1-z} E_{sp}(R, \mathcal{W}) \lambda(dz) + \frac{1}{\epsilon} \int_{\alpha - \epsilon\alpha}^{\alpha} E_{sp}(R, \mathcal{W}) \lambda(dz) + \frac{\epsilon}{1-\epsilon} R \\ &\leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{1-\epsilon} \frac{1-\alpha}{\alpha} E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{1-\epsilon} R \end{aligned}$$

Since $(1 - \alpha) E_{sp}(R, \mathcal{W}) + \alpha R \leq (R \vee E_{sp}(R, \mathcal{W}))$ for all $\alpha \in (0, 1)$ we have,

$$\frac{1-\alpha}{\alpha} (\tilde{C}_{\alpha, \mathcal{W}}^\epsilon - R) \leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{\phi - \epsilon} (R \vee E_{sp}(R, \mathcal{W})) \quad \alpha \in [\frac{\phi - \epsilon}{1 - \epsilon}, 1). \quad (53)$$

Then as a result of equations (52) and (53) we have

$$\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{\phi - \epsilon} (R \vee E_{sp}(R, \mathcal{W})) \quad \forall R \in [C_{\phi, \mathcal{W}}, C_{1, \mathcal{W}}]. \quad (54)$$

$C_{\alpha, \mathcal{W}}$ is increasing in α by [35, Lemma 11-(a)] and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is decreasing in α on $(0, 1)$ by [35, Lemma 11-(c)]. Then as a result of the expression for $E_{sp}(R, \mathcal{W})$ given in equation (12), we have $E_{sp}(R, \mathcal{W}) \leq \frac{1-\phi}{\phi} R$ for all $R \in [C_{\phi, \mathcal{W}}, C_{1, \mathcal{W}}]$. Hence

$$\frac{\epsilon}{\phi - \epsilon} (R \vee E_{sp}(R, \mathcal{W})) \leq \frac{\epsilon}{\phi - \epsilon} \frac{R}{\phi} \quad \forall R \in [C_{\phi, \mathcal{W}}, C_{1, \mathcal{W}}]. \quad (55)$$

Then (48) follows from (50), (54) and (55).

In order to prove (49), first note that $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is decreasing in R by definition. Then for any \mathcal{W} with finite $C_{1, \mathcal{W}}$ we have

$$\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \leq \inf_{x \in (0, C_{1, \mathcal{W}})} \tilde{E}_{sp}^\epsilon(x, \mathcal{W}) \quad \forall R \in [C_{1, \mathcal{W}}, \infty).$$

Then (49) follows from (48) and Lemma 3 (i.e. $E_{sp}(C_{1, \mathcal{W}}, \mathcal{W}) = 0$ and the continuity of $E_{sp}(R, \mathcal{W})$ in R). ■

D. Outer Bounds for Codes on Product Channels

In this subsection we derive outer bounds for codes on product channels using the lower bound on small deviation probability given in Lemma 9, the intermediate value theorem and a pigeon hole argument. First, we establish a bound in terms of the averaged sphere packing exponent that is valid for all product channels, see Lemma 12. Then we sharpen the bound for product channels whose Renyi centers do not change with the order, see Lemma 13. After that we point out that for certain product channels with feedback the independence hypothesis of Lemma 9 is satisfied and the bound given in Lemma 13 for the corresponding product channels holds as is, see Lemma 14.

Lemma 12. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a product channel for the index set $\mathcal{T} = \{1, \dots, n\}$ and M, L be integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{\tilde{C}_{\alpha_0, \mathcal{W}} + \frac{\gamma_{\kappa, \epsilon}}{1-\alpha_0}}$ for a $\kappa \geq 3$, an $\alpha_0 \in (0, 1)$ and an $\epsilon \in (0, 1)$ satisfying $\frac{(n-1)(1-\alpha_0)(1-\epsilon)}{\epsilon} \geq 1$. Then any (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ satisfies*

$$P_e^{av} \geq \left(\frac{\epsilon e^{-2\gamma_{\kappa, \epsilon}}}{16e^2(1-\alpha_0)n^{3/2}} \right)^{\frac{1}{\alpha_0}} e^{-\tilde{E}_{sp}^\epsilon(R, \mathcal{W})} \quad R = \ln \frac{M}{L} \quad (56)$$

where $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is defined in equation (47) and $\gamma_{\kappa, \epsilon}$ is given by

$$\gamma_{\kappa, \epsilon} \triangleq \frac{3\sqrt[3]{3}}{1-\epsilon} \left(\sum_{t=1}^n (C_{1/2, \mathcal{W}_t} \vee \kappa)^\kappa \right)^{\frac{1}{\kappa}}. \quad (57)$$

We have presented Lemma 12 using the bound given in (56) in order to emphasize the similarity to the Gallager's inner bound given in Lemma 4. However, bound given in (56) becomes trivial as α_0 converges to zero. By changing the analysis slightly it is possible to obtain an alternative bound given in (58). The bound given in (58) is preferable especially for codes with low rates on channels satisfying $\lim_{R \downarrow 0} E_{sp}(R, \mathcal{W}) < \infty$.

$$P_e^{av} \geq \left(\frac{\epsilon}{16(1-\alpha_0)n^{3/2}} \right) e^{-2\gamma_{\kappa, \epsilon}} e^{-\tilde{E}_{sp}^\epsilon(R, \mathcal{W})} \quad R = \ln \frac{M}{L} - 2\gamma_{\kappa, \epsilon} - \ln \frac{16e^2(1-\alpha_0)n^{3/2}}{\epsilon}. \quad (58)$$

Proof of Lemma 12 and equation (58): Let (Ψ, Θ) be a (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$. We denote $(\mathcal{Y}_t, \mathcal{Y}_t)$ marginal of $\Psi(m)$ by $\Psi_t(m)$. Let $v_{\alpha, t}^m$ be a probability measure on $(\mathcal{Y}_t, \mathcal{Y}_t)$ defined through its Radon-Nikodym derivative:

$$\frac{dv_{\alpha, t}^m}{d\nu} \triangleq e^{(1-\alpha)D_\alpha(\Psi_t(m) \| q_{\alpha, \mathcal{W}_t}^\epsilon)} \left(\frac{d\Psi_t(m)}{d\nu} \right)^\alpha \left(\frac{dq_{\alpha, \mathcal{W}_t}^\epsilon}{d\nu} \right)^{1-\alpha}$$

where $q_{\alpha, \mathcal{W}_t}^\epsilon$ is the average Renyi center of \mathcal{W}_t defined in (41) and ν is any reference measure for $\Psi_t(m)$ and $q_{\alpha, \mathcal{W}_t}^\epsilon$. Let $X_{\alpha, t}^m$ be a random variable defined as

$$X_{\alpha, t}^m \triangleq \ln \frac{d(\Psi_t(m))_\sim}{dq_{\alpha, \mathcal{W}_t}^\epsilon} - \mathbf{E}_{v_{\alpha, t}^m} \left[\ln \frac{d(\Psi_t(m))_\sim}{dq_{\alpha, \mathcal{W}_t}^\epsilon} \right]$$

where $(\Psi_t(m))_\sim$ is the component of $\Psi_t(m)$ that is absolutely continuous in $q_{\alpha, \mathcal{W}_t}^\epsilon$.

Note that $X_{\alpha, t}^m$ can be written in terms of $\ln \frac{dv_{\alpha, t}^m}{d\Psi_t(m)}$ or $\ln \frac{dv_{\alpha, t}^m}{dq_{\alpha, \mathcal{W}_t}^\epsilon}$:

$$X_{\alpha, t}^m = \frac{1}{\alpha-1} \left(\ln \frac{dv_{\alpha, t}^m}{d\Psi_t(m)} - D_1(v_{\alpha, t}^m \| \Psi_t(m)) \right) = \frac{1}{\alpha} \left(\ln \frac{dv_{\alpha, t}^m}{dq_{\alpha, \mathcal{W}_t}^\epsilon} - D_1(v_{\alpha, t}^m \| q_{\alpha, \mathcal{W}_t}^\epsilon) \right).$$

Let the probability measures q_α^ϵ and v_α^m on $(\mathcal{Y}, \mathcal{Y})$ and the random variable X_α^m in the probability space $(\mathcal{Y}, \mathcal{Y}, v_\alpha^m)$ be

$$q_\alpha^\epsilon \triangleq \prod_{t \in \mathcal{T}}^\otimes q_{\alpha, \mathcal{W}_t}^\epsilon \quad v_\alpha^m \triangleq \prod_{t \in \mathcal{T}}^\otimes v_{\alpha, t}^m \quad X_\alpha^m \triangleq \sum_{t \in \mathcal{T}} X_{\alpha, t}^m.$$

As a result of the product structure of the probability measures $\Psi(m)$, q_α^ϵ and v_α^m we have

$$D_\alpha(\Psi(m) \| q_\alpha^\epsilon) = \sum_{t=1}^n D_\alpha(\Psi_t(m) \| q_{\alpha, \mathcal{W}_t}^\epsilon) \quad (59)$$

$$X_\alpha^m = \frac{1}{\alpha-1} \left(\ln \frac{dv_\alpha^m}{d\Psi(m)} - D_1(v_\alpha^m \| \Psi(m)) \right) = \frac{1}{\alpha} \left(\ln \frac{dv_\alpha^m}{dq_\alpha^\epsilon} - D_1(v_\alpha^m \| q_\alpha^\epsilon) \right). \quad (60)$$

Bounding each term in the sum via Lemma 10 and using equation (44) we get

$$D_\alpha(\Psi(m) \| q_\alpha^\epsilon) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon. \quad (61)$$

Let $\mathcal{E}_m \in \mathcal{Y}$ be $\mathcal{E}_m \triangleq \{y : m \in \Theta(y)\}$. Then for any two real numbers τ_1 and τ_2 we have

$$\begin{aligned} P_e^m &\geq e^{-D_1(v_\alpha^m \| \Psi(m)) - \tau_1} \int_{\mathcal{Y} \setminus \mathcal{E}_m} \mathbf{1}_{\{\ln \frac{dv_\alpha^m}{dq_\alpha^\epsilon} \leq D_1(v_\alpha^m \| \Psi(m)) + \tau_1\}} v_\alpha^m(dy) \\ q_\alpha^\epsilon(\mathcal{E}_m) &\geq e^{-D_1(v_\alpha^m \| q_\alpha^\epsilon) - \tau_2} \int_{\mathcal{E}_m} \mathbf{1}_{\{\ln \frac{dv_\alpha^m}{dq_\alpha^\epsilon} \leq D_1(v_\alpha^m \| q_\alpha^\epsilon) + \tau_2\}} v_\alpha^m(dy) \\ P_e^m e^{D_1(v_\alpha^m \| \Psi(m)) + \tau_1} + q_\alpha^\epsilon(\mathcal{E}_m) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \tau_2} &\geq \int_{\frac{\tau_1}{\alpha-1} \leq X_\alpha^m \leq \frac{\tau_2}{\alpha}} v_\alpha^m(dy). \end{aligned}$$

Random variables $X_{\alpha,t}^m$ for $t \in \mathcal{T}$ are zero mean in the probability space $(\mathcal{Y}, \mathcal{Y}, v_\alpha^m)$ by construction. Furthermore, they are jointly independent because of the product structure of the probability measures $\Psi(m)$, q_α^ϵ and v_α^m . Applying³⁰ Lemma 9 for $\tau_1 = (1 - \alpha)\xi_{\alpha,\kappa}^m$ and $\tau_2 = \alpha\xi_{\alpha,\kappa}^m$ we get,

$$P_e^m e^{D_1(v_\alpha^m \|\Psi(m)) + (1-\alpha)\xi_{\alpha,\kappa}^m} + q_\alpha^\epsilon(\mathcal{E}_m) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \alpha\xi_{\alpha,\kappa}^m} \geq \frac{1}{2\sqrt{n}} \quad (62)$$

where

$$\xi_{\alpha,\kappa}^m \triangleq 3 \left(\sum_{t=1}^n \mathbf{E}_{v_\alpha^m} [|X_{\alpha,t}^m|^\kappa] \right)^{\frac{1}{\kappa}}. \quad (63)$$

Note that $\mathbf{E}_{v_\alpha^m} [|X_{\alpha,t}^m|^\kappa] = \mathbf{E}_{v_{\alpha,t}^m} [|X_{\alpha,t}^m|^\kappa]$ by construction. Then using Lemma 6 get,

$$\mathbf{E}_{v_\alpha^m} [|X_{\alpha,t}^m|^\kappa] \leq 3 \left(\frac{((1-\alpha)D_\alpha(\Psi_t(m) \| q_{\alpha,w_t}^\epsilon)) \vee \kappa}{\alpha(1-\alpha)} \right)^\kappa \quad (64)$$

We can bound $D_\alpha(\Psi_t(m) \| q_{\alpha,w_t}^\epsilon)$ using Lemma 10

$$(1 - \alpha)D_\alpha(\Psi_t(m) \| q_{\alpha,w_t}^\epsilon) \leq \frac{C_{1/2, w_t}}{1 - \epsilon}. \quad (65)$$

Using the definitions of $\gamma_{\kappa,\epsilon}$ and $\xi_{\alpha,\kappa}^m$ given in equations (57) and (63) together with the equations (64) and (65) we get

$$\xi_{\alpha,\kappa}^m \leq \frac{\gamma_{\kappa,\epsilon}}{\alpha(1-\alpha)}. \quad (66)$$

In the following, we describe a subset of the message set with at least $\approx \frac{M\epsilon}{n(1-\alpha_0)}$ messages such that P_e^m is at least $\approx \frac{e^{-\tilde{E}_{sp}^\epsilon(R, \mathcal{W}) - 2\gamma_{\kappa,\epsilon}}}{\sqrt{n}}$ for all m in the subset. We prove the inequality given in (56) using this subset together with equations (62) and (66). Let us consider the subset of the message set, \mathcal{M}_1 defined as follows:

$$\begin{aligned} \mathcal{M}_1 &\triangleq \left\{ m : (q_\alpha^\epsilon(\mathcal{E}_m) + \frac{(1-\alpha_0)L}{M}) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} \geq \frac{1}{4\sqrt{n}} \quad \forall \alpha \in (\alpha_0, 1) \right\} \\ \mathcal{M} \setminus \mathcal{M}_1 &= \left\{ m : \exists \alpha \in (\alpha_0, 1) \text{ such that } (q_\alpha^\epsilon(\mathcal{E}_m) + \frac{(1-\alpha_0)L}{M}) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} < \frac{1}{4\sqrt{n}} \right\} \end{aligned}$$

First, we bound the size of \mathcal{M}_1 from above. We can bound $D_1(v_\alpha^m \| q_\alpha^\epsilon)$ using the non-negativity of the Renyi divergence for probability measures, i.e. [35, Lemma 9-(g)], the definitions of v_α^m and $v_{\alpha,t}^m$'s, and equation (61)

$$\begin{aligned} D_1(v_\alpha^m \| q_\alpha^\epsilon) &= D_\alpha(\Psi(m) \| q_\alpha^\epsilon) - \frac{\alpha}{1-\alpha} D_1(v_\alpha^m \| \Psi(m)) \\ &\leq \tilde{C}_{\alpha,\mathcal{W}}^\epsilon. \end{aligned} \quad \forall m \in \mathcal{M}, \alpha \in (0, 1) \quad (67)$$

Then summing the inequality in the condition for membership in \mathcal{M}_1 over the members of \mathcal{M}_1 we get

$$2Le^{\tilde{C}_{\alpha,\mathcal{W}}^\epsilon + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} \geq |\mathcal{M}_1| \frac{1}{4\sqrt{n}} \quad \forall \alpha \in (\alpha_0, 1) \quad \Rightarrow \quad \frac{|\mathcal{M}_1|}{L} \leq 8\sqrt{n} e^{\tilde{C}_{\alpha_0,\mathcal{W}}^\epsilon + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_0}}.$$

Consequently $|\mathcal{M}_1| < \frac{M}{2}$ because $\frac{M}{L} > 16\sqrt{n} e^{\tilde{C}_{\alpha_0,\mathcal{W}}^\epsilon + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_0}}$ by hypothesis.

On the other hand, as a result of definition of \mathcal{M}_1 , for each $m \in \mathcal{M} \setminus \mathcal{M}_1$ there is an α satisfying

$$(q_\alpha^\epsilon(\mathcal{E}_m) + \frac{(1-\alpha_0)L}{M}) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} < \frac{1}{4\sqrt{n}}.$$

Furthermore, $q_\alpha^\epsilon(\mathcal{E}_m)$ is continuous in α because $|q_\alpha^\epsilon(\mathcal{E}_m) - q_z^\epsilon(\mathcal{E}_m)| \leq \frac{|\alpha - z|}{\epsilon}$ and $D_1(v_\alpha^m \| q_\alpha^\epsilon)$ is continuous in α by Lemma 7. Thus $(q_\alpha^\epsilon(\mathcal{E}_m) + \frac{(1-\alpha_0)L}{M}) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}}$ is continuous in α . Since $\lim_{\alpha \uparrow 1} (q_\alpha^\epsilon(\mathcal{E}_m) + \frac{L}{M}) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} = \infty$, using the intermediate value theorem [42, 4.23] we can conclude that for each $m \in \mathcal{M} \setminus \mathcal{M}_1$ there exists an $\alpha_m \in (\alpha_0, 1)$ such that

$$(q_{\alpha_m}^\epsilon(\mathcal{E}_m) + \frac{(1-\alpha_0)L}{M}) e^{D_1(v_{\alpha_m}^m \| q_{\alpha_m}^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_m}} = \frac{1}{4\sqrt{n}}. \quad (68)$$

There exists a length $\frac{1-\alpha_0}{K}$ interval with $\lceil \frac{M}{2K} \rceil$ or more α_m 's. Let $[z, z + \frac{1-\alpha_0}{K}]$ be the aforementioned interval. Using the definition of the averaged Renyi center given in (41) and choosing $\tilde{\epsilon}$ and $\tilde{\alpha}$ to be $\tilde{\epsilon} = \frac{1-\alpha_0}{K} + \epsilon(1 - \frac{1-\alpha_0}{K})$ and $\tilde{\alpha} = \frac{1-\epsilon}{1-\tilde{\epsilon}}z$, respectively, we get

$$q_{\alpha,w_i}^\epsilon \leq \tilde{\epsilon} q_{\tilde{\alpha},w_i}^\epsilon \quad q_\alpha^\epsilon \leq (\tilde{\epsilon})^n q_{\tilde{\alpha}}^\epsilon \quad \forall \alpha \in [z, z + \frac{1-\alpha_0}{K}].$$

At least half of the messages with α_m 's in $[z, z + \frac{1-\alpha_0}{K}]$, i.e. at least $\lceil \frac{1}{2} \lceil \frac{M}{2K} \rceil \rceil$ messages, satisfy

$$q_{\tilde{\alpha}}^\epsilon(\mathcal{E}_m) \leq 2 \frac{L}{\lceil \frac{M}{2K} \rceil} \leq \frac{4L}{M} K \quad \Rightarrow \quad q_\alpha^\epsilon(\mathcal{E}_m) \leq \frac{4L}{M} K \left(1 + \frac{1-\alpha_0}{K} \frac{1-\epsilon}{\epsilon} \right)^n$$

³⁰We can apply Chebyshev inequality and obtain the bound $P_e^m e^{D_1(v_\alpha^m \|\Psi(m)) + (1-\alpha)\xi_{\alpha,2}^m} + q_\alpha^\epsilon(\mathcal{E}_m) e^{D_1(v_\alpha^m \| q_\alpha^\epsilon) + \alpha\xi_{\alpha,2}^m} \geq \frac{8}{9}$.

If we set K to be $K = \lfloor \frac{(n-1)(1-\alpha_0)(1-\epsilon)}{\epsilon} \rfloor$ and use the identity $(1+x)^{\frac{1}{x}} < e$ we get³¹

$$q_\alpha^\epsilon(\mathcal{E}_m) \leq \frac{4L}{M} \frac{n(1-\alpha_0)(1-\epsilon)}{\epsilon} \left(1 + \frac{(1-\alpha_0)(1-\epsilon)}{K\epsilon}\right)^{n-1} \leq \frac{4L}{M} \frac{n(1-\alpha_0)(1-\epsilon)}{\epsilon} e^2. \quad (69)$$

Then using (68) we can bound $D_1(v_{\alpha_m}^m \| q_{\alpha_m}^\epsilon)$ for all m satisfying (69)

$$e^{D_1(v_{\alpha_m}^m \| q_{\alpha_m}^\epsilon) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_m}} \geq \frac{1}{4\sqrt{n}} \frac{\epsilon}{4e^2(1-\alpha_0)n} \frac{M}{L}. \quad (70)$$

On the other hand we can bound P_e^m using (62), (66) and (68)

$$P_e^m e^{D_1(v_{\alpha_m}^m \| \Psi(m)) + \frac{\gamma_{\kappa,\epsilon}}{\alpha_m}} \geq \frac{1}{4\sqrt{n}}. \quad (71)$$

Using equations (61), (67), (70) and (71) we get

$$P_e^m e^{\frac{1-\alpha_m}{\alpha_m} \tilde{C}_{\alpha_m, W}^\epsilon + 2\frac{\gamma_{\kappa,\epsilon}}{\alpha_m}} \geq \left(\frac{1}{4\sqrt{n}}\right) \left(\frac{1}{4\sqrt{n}} \frac{\epsilon}{4e^2(1-\alpha_0)n} \frac{M}{L}\right)^{\frac{1-\alpha_m}{\alpha_m}}. \quad (72)$$

Hence, for all m satisfying (69) as a result of definition of $\tilde{E}_{sp}^\epsilon(R, W)$ given in (47) we have

$$P_e^m \geq e^{-2\frac{\gamma_{\kappa,\epsilon}}{\alpha_0}} \left(\frac{1}{4\sqrt{n}}\right) \left(\frac{1}{4\sqrt{n}} \frac{\epsilon}{4e^2(1-\alpha_0)n}\right)^{\frac{1-\alpha_0}{\alpha_0}} e^{-\tilde{E}_{sp}^\epsilon(R, W)} \quad R = \ln \frac{M}{L}.$$

Since there are at least $\lceil \frac{M}{4K} \rceil$ such messages we get the inequality in (56).

Note that $\left(\frac{\epsilon e^{-2\gamma_{\kappa,\epsilon}}}{16e^2(1-\alpha_0)n^{3/2}}\right)^{\frac{1}{\alpha_0}}$ diverges as α_0 converge to zero. In order avoid this phenomena, one can change the analysis after equation (72) and introduce an approximation error term to the rate of the averaged sphere packing exponent term: For all for all m satisfying (69) as a result of the definition of $\tilde{E}_{sp}^\epsilon(R, W)$ given in (47) and equation (72) we have

$$P_e^m \geq \left(\frac{1}{4\sqrt{n}}\right) e^{-2\gamma_{\kappa,\epsilon}} e^{-\tilde{E}_{sp}^\epsilon(R, W)} \quad R = \ln \frac{M}{L} - 2\gamma_{\kappa,\epsilon} - \ln \frac{16e^2(1-\alpha_0)n^{3/2}}{\epsilon}.$$

Since there are at least $\lceil \frac{M}{4K} \rceil$ such messages we get the inequality given in (58). ■

For certain channels the Renyi center does not change with the order on the interval it exits, or on a subset of it. In those cases one can improve Lemma 12 as follows:

Lemma 13. *Let $((\mathcal{Y}, \mathcal{Y}), W)$ be a product channel for the index set $\mathcal{T} = \{1, \dots, n\}$ satisfying*

$$q_{\alpha, W_t} = q_t \quad \forall \alpha \in [\alpha_0, 1) \quad (73)$$

for a $q_t \in \mathcal{P}(\mathcal{Y}_t, \mathcal{Y}_t)$ for each $t \in \mathcal{T}$ and M, L be integers satisfying $\frac{M}{L} > 16\sqrt{n} e^{C_{\alpha_0, W} + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_0}}$ for a $\kappa \geq 3$. Then any (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), W)$ satisfies

$$P_e^{av} \geq \left(\frac{e^{-2\gamma_{\kappa,\epsilon}}}{8n^{1/2}}\right) e^{-E_{sp}(R, W)} \quad R = \ln \frac{M}{L} - 2\gamma_{\kappa,\epsilon} - \frac{\ln n}{2} - \ln 8 \quad (74)$$

where $E_{sp}(R, W)$ is defined in equation (11) and γ_{κ} is given by

$$\gamma_{\kappa} = 3\sqrt[3]{3} \left(\sum_{t=1}^n (C_{1/2, W_t} \vee \kappa)^\kappa\right)^{\frac{1}{\kappa}}. \quad (75)$$

Any product of shift invariant channels, described in [35, Example 7], satisfies the constraint given in equation (73) because $q_{\alpha, W^{[\mathcal{T}]}} = \lambda$ if $q_{\alpha, W^{[t]}}$ exists.

The Renyi center of a product channel is the product of the Renyi centers of the component channels, by Lemma 1. Thus the hypothesis of Lemma 13 is equivalent to the constraint

$$q_{\alpha, W} = q_{\alpha_0, W} \quad \forall \alpha \in (\alpha_0, 1).$$

Proof of Lemma 13: We use q_t 's rather than q_{α, W_t}^ϵ 's to define the probability measure q_α on $(\mathcal{Y}_{[1,n]}, \mathcal{Y}_{[1,n]})$ and we define \mathcal{M}_1 as

$$\mathcal{M}_1 \triangleq \left\{ m : \left(q_\alpha(\mathcal{E}_m) + \frac{L}{M}\right) e^{D_1(v_\alpha^m \| q_\alpha) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha}} \geq \frac{1}{4\sqrt{n}} \quad \forall \alpha \in (\alpha_0, 1) \right\}. \quad (76)$$

We repeat the analysis we have done for Lemma 12 up to equation (68):

$$\left(q_{\alpha_m}(\mathcal{E}_m) + \frac{L}{M}\right) e^{D_1(v_{\alpha_m}^m \| q_{\alpha_m}) + \frac{\gamma_{\kappa,\epsilon}}{1-\alpha_m}} = \frac{1}{4\sqrt{n}}. \quad (77)$$

There exists at least $\lceil \frac{M}{2} \rceil$ messages with

$$q_{\alpha_m}(\mathcal{E}_m) \leq 2\frac{L}{M} \quad (78)$$

³¹We have also assumed $K \geq 1$ and this assumption leads to the condition $\frac{(n-1)(1-\alpha_0)(1-\epsilon)}{\epsilon} \geq 1$.

Using (78) instead of (69) and repeating the rest of the analysis we get (74). \blacksquare

The small deviations bound given in Lemma 9 is another key ingredient of the proof of Lemma 12. The independence hypothesis of Lemma 9 is implied by the product structure of each $w \in \mathcal{W}$ and $q_{\alpha, \mathcal{W}}$. However, the product structure is not necessary for the independence, provided that the channel has certain symmetries.

Lemma 14. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a product channel with feedback for the index set $\mathcal{T} = \{1, \dots, n\}$ satisfying³²*

$$q_{\alpha, \mathcal{W}_t} = q_t \quad \forall \alpha \in (0, 1) \quad (79)$$

$$q_t(\frac{dw_{\sim}}{dq_t} \leq x) = g_t(x) \quad \forall w \in \mathcal{W}_t \quad (80)$$

for a $q_t \in \mathcal{P}(\mathcal{Y}_t, \mathcal{Y}_t)$ and a cumulative distribution function $g_t : [0, \infty) \rightarrow [0, 1]$ for each $t \in \mathcal{T}$ and M, L be integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{C_{\alpha_0, \mathcal{W}} + \frac{\gamma_{\kappa}}{1-\alpha_0}}$ for a $\kappa \geq 3$. Then any (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ satisfies the equation (74) for γ_{κ} given in equation (75).

Any product channel whose component channels are modular shift channels described in [35, Example 4], satisfy the constraints given in (79) and (80). Products of more general shift invariant channels described in [35, Example 7], do satisfy the constraint given in (79) but they may or may not satisfy the constraint given in (80) depending on \mathcal{F} .

Proof of Lemma 14: As we have done for Lemma 13 we use q_t 's rather than $q_{\alpha, \mathcal{W}_t}^{\epsilon}$'s and to define the product probability measure q_{α} on $(\mathcal{Y}_{[1, n]}, \mathcal{Y}_{[1, n]})$. Since we have a product channel with feedback $\Psi(m)$ is not necessarily a product measure. However, as a result of hypothesis of the lemma given in (80), $X_{\alpha, t}^m$'s for $t \in \mathcal{T}$ are jointly independent random variables in the probability space $(\mathcal{Y}, \mathcal{Y}, \nu_{\alpha}^m)$ for any $\alpha \in (0, 1)$ and $m \in \mathcal{M}$. Rest of the proof is identical to the proof of Lemma 13. \blacksquare

E. Proof of Theorem 2

Proof of Theorem 2: We prove Theorem 2 using Lemmas 11 and 12. We are free to choose different values for ϵ and κ for different values of n , provided that the hypotheses of Lemmas 11 and 12 are satisfied.

As a result of Assumption 1 there exists a $K \in [1, \infty)$, $n_0 \in \mathbb{Z}^+$ such that

$$\max_{t \in [1, n]} C_{1/2, \mathcal{W}_t} \leq K \ln n \quad \forall n \geq n_0. \quad (81)$$

Let κ_n be $\kappa_n = K \ln n$ and ϵ_n be any $o(1)$ function. Then for n large enough

$$\gamma_{\kappa_n, \epsilon_n} \leq 4eK \ln n. \quad (82)$$

$C_{\alpha, \mathcal{W}_{[1, n]}}$ is increasing in α by [35, Lemma 11-(a)] and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}_{[1, n]}}$ is decreasing in α on $(0, 1)$ by [35, Lemma 11-(c)]. Thus, we can bound $\tilde{C}_{\alpha, \mathcal{W}_{[1, n]}}^{\epsilon}$ using the definition of averaged Renyi capacity given in equation (43):

$$\tilde{C}_{\alpha, \mathcal{W}_{[1, n]}}^{\epsilon} \leq \left(1 + \frac{\epsilon}{1-\epsilon} \frac{\alpha^2 + (1-\alpha)^2}{\alpha(1-\alpha)}\right) C_{\alpha, \mathcal{W}_{[1, n]}}.$$

Then for $\epsilon_n = \frac{1}{n}$ and for n large enough by equations (81) and (82) we have

$$16\sqrt{n}e^{\tilde{C}_{\alpha_0, \mathcal{W}_{[1, n]}}^{\epsilon} + \frac{\gamma_{\kappa, \epsilon}}{1-\alpha_0}} \leq 16e^{C_{\alpha_0, \mathcal{W}_{[1, n]}} + (\frac{1}{2} + \frac{4eK}{1-\alpha_0} + \frac{K}{\alpha_0(1-\alpha_0)}) \ln n}. \quad (83)$$

Thus as a result of the hypothesis of the theorem, hypotheses of Lemma 12 is satisfied for all n large enough. Thus using (82) we can conclude that for n large enough

$$P_e^{av} \geq \left(\frac{\epsilon_n n^{-8eK}}{16e^2(1-\alpha_0)n^{3/2}}\right)^{\frac{1}{\alpha_0}} e^{-\tilde{E}_{sp}^{\epsilon_n}(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]})} \quad \text{where } \epsilon_n = \frac{1}{n}. \quad (84)$$

On the other hand Lemma 11, the hypothesis of the theorem given in (23), and the monotonicity $C_{\alpha, \mathcal{W}}$ in α imply that for n large enough

$$\tilde{E}_{sp}^{\epsilon_n} \left(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]} \right) \leq E_{sp} \left(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]} \right) + \frac{C_{\alpha_1, \mathcal{W}_{[1, n]}}}{(\alpha_0 - 1/n)\alpha_0}.$$

Using equation (81) and the monotonicity of $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ and $C_{\alpha, \mathcal{W}}$ in α we can conclude that for n large enough

$$\tilde{E}_{sp}^{\epsilon_n} \left(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]} \right) \leq E_{sp} \left(\ln \frac{M_n}{L_n}, \mathcal{W}_{[1, n]} \right) + \frac{1}{(\alpha_0 - 1/n)\alpha_0} \left(\frac{\alpha_1}{1-\alpha_1} \vee 1 \right) K \ln n. \quad (85)$$

Then equation (24) follows from (84) and (85). \blacksquare

³² w_{\sim} stands for the component of w that is absolutely continuous in q_t .

IV. THE SPHERE PACKING BOUND FOR PRODUCT CHANNELS WITH FEEDBACK

For certain product channels with feedback, Lemma 14 establishes an outer bound that is equivalent to the one established by Lemma 12 for product channels. If every channel in the sequence $\{\mathcal{W}_{[1,n]}\}_{n \in \mathbb{Z}^+}$ satisfies the hypothesis of Lemma 14 then we can use Lemma 14 to prove Theorem 2, instead of Lemma 12. Hence, for codes on such a sequence of product channels with feedback, Theorem 2 holds as is. First Dobrushin [19] and then Haroutunian [29] employed a similar observation to claim the validity of the sphere packing bound for certain DSPCs with feedback. Later, Augustin [9, p318] did the same for certain product channels with feedback.

For arbitrary product channels with feedback, however, we do not have a parametric outer bound similar to the one established in Lemma 12, yet. We can, however, recover a somewhat weaker parametric outer bound given in Lemma 17 by assuming stationarity and discreteness. Theorem 3 is the corresponding asymptotic result.

Theorem 3. *Let $\{\mathcal{W}_t\}_{t \in \mathbb{Z}^+}$ be a stationary sequence of discrete channels such that $\mathcal{W}_t = \mathcal{W}_0$ for all $t \in \mathbb{Z}^+$ and α_0, α_1 be orders satisfying $0 < \alpha_0 < \alpha_1 < 1$. Then for any sequence of codes on product channels with feedback $\{\mathcal{W}_{[1,n]}\}_{n \in \mathbb{Z}^+}$ satisfying*

$$nC_{\alpha_1, \mathcal{W}_0} \geq \ln \frac{M_n}{L_n} \geq nC_{\alpha_0, \mathcal{W}_0} + 4n^{\frac{3}{4}} \ln n \quad \forall n \geq n_0 \quad (86)$$

there exists a $\tau \geq \mathbb{R}^+$ and an $n_1 \geq n_0$ such that

$$P_e^{av} \geq e^{-nE_{sp}(\frac{1}{n} \ln \frac{M_n}{L_n}, \mathcal{W}_0) - \tau n^{\frac{3}{4}} \ln n} \quad \forall n \geq n_1. \quad (87)$$

We use Augustin's method together with the ideas from Sheverdyaev [48] and Haroutunian [29] to prove Theorem 3. In Section IV-A, we establish a Taylor's expansion for $D_\alpha(w \| q)$ around $\alpha = 1$ assuming $D_\phi(w \| q)$ is finite for a $\phi > 1$. In Section IV-B, we first present a brief review of the dummy channel method for establishing outer bounds and then we prove that for every $R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$ there exists a dummy channel $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ and an $x \in (1, \infty)$ such that $C_{x, f(\mathcal{W})} \lesssim R$ and $\sup_{w \in \mathcal{W}} D_1(f(w) \| w) \lesssim E_{sp}(R, \mathcal{W})$. In Section IV-C, we derive a parametric outer bound using a variant of Haroutunian's technique, [28], [29]. We use this outer bound to prove Theorem 3 in Section IV-D.

Before starting the proof of Theorem 3, we make a brief digression to discuss possible extensions of Theorem 3 and to compare it with certain related results. The stationarity assumption of Theorem 3 can be relaxed slightly without making any major changes in the proofs. In particular, one can prove Theorem 3 for any sequence of discrete channels satisfying Assumption 4 given in the following.

Assumption 4. $\{\mathcal{W}_t\}_{t \in \mathbb{N}}$ is an ordered sequence of channels such that for any $\alpha_0 \in (0, 1)$ and $\alpha_1 \in (\alpha_0, 1)$, the difference $\sup_{\alpha \in [\alpha_0, \alpha_1]} \sup_{t \in \mathbb{N}} |C_{\alpha, \mathcal{W}_{(t, t+n]}} - nC_{\alpha, \mathcal{W}_0}|$ is $o(n^{\frac{3}{4}} \ln n)$, i.e.

$$\forall \alpha_0 \in (0, 1), \alpha_1 \in (\alpha_0, 1), \varepsilon \in \mathbb{R}^+, \exists n_0 \text{ such that } \sup_{\alpha \in [\alpha_0, \alpha_1]} \sup_{t \in \mathbb{N}} |C_{\alpha, \mathcal{W}_{(t, t+n]}} - nC_{\alpha, \mathcal{W}_0}| \leq \varepsilon n^{\frac{3}{4}} \ln n \text{ for all } n \geq n_0.$$

We have confined the claims of Theorem 3 to discrete channels in order avoid certain measurability issues. We believe, however, it should be possible to resolve those issues and to extend Theorem 3 to any sequence of channels satisfying Assumption 4 for a \mathcal{W}_0 satisfying $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}_0} = 0$. Augustin makes the same conjecture for the stationary channels in [9, Corollary 41.9]. If valid, such an extension of Theorem 3 will imply the sphere packing bound for the Poisson channel $\Lambda^{[T, a, b]}$ described in [35, equation (103)] even when it is extended by the inclusion of the intensity functions whose value at any time τ depends on the arrivals in the interval $[0, \tau - \epsilon]$ for some fixed $\epsilon > 0$. As it was the case for the results about channel capacity, discussed on page 11, the extension to zero-delay case requires application of martingale arguments.

Augustin presents a proof sketch for a result, [9, Theorem 41.7], slightly stronger than Theorem 3. The approximation error terms in [9, Theorem 41.7] are $O(n^{\frac{2}{3}} \ln n)$ rather than $O(n^{\frac{3}{4}} \ln n)$. Augustin [9] claims his sketch works for arbitrary output spaces $(\mathcal{Y}_0, \mathcal{Y}_0)$ and finite \mathcal{W}_0 ; we can confirm Augustin's proof sketch only for discrete channels. Augustin's proof sketch is effected by measurability issues similar to the ones mentioned in the previous paragraph.

A. A Taylor's Expansion For The Renyi Divergence

Sheverdyaev [48] used Taylor's expansion to prove a result equivalent to Theorem 3. However, the approximation error terms are not explicit in [48]. Recently, Fong and Tan [23, Proposition 11] bounded $D_z(w \| q)$ for $z \in [1, \frac{5}{4}]$ using Taylor's expansion for the case when \mathcal{Y} is a finite set and \mathcal{Y} is $\mathcal{Z}^{\mathcal{Y}}$. Fong and Tan's bound [23, Proposition 11], however, is not appropriate for our purposes because the approximation error is proportional to $|\mathcal{Y}|$. Assuming finite $D_\infty(w \| q)$, i.e. bounded $\frac{dw}{dq}$, Sason and Verdú derived a similar bound [43, Theorem 35-(b), (469)].³³ In the following we bound $D_z(w \| q)$ for $z \in (1, \phi)$ using Taylor's expansion assuming $D_\phi(w \| q)$ is finite.

³³Guntuboyina, Saha and Schiebinger [26] established a general method for establishing sharp bounds among f -divergences, without assuming either $D_\infty(w \| q)$ or $D_\infty(q \| w)$ to be finite. Yet such conditions can easily be included in the framework proposed in [26].

Lemma 15. Let w and q be two probability measures on the measurable space $(\mathcal{Y}, \mathcal{Y})$ satisfying $D_\phi(w \| q) \leq \gamma$ for a $\gamma \in (0, \infty)$ and a $\phi \in (1, \infty)$. Then for any $z \in (1, \phi)$

$$0 \leq D_z(w \| q) - D_1(w \| q) \leq \inf_{\alpha \in (z, \phi]} \frac{4(z-1)}{e^2(\alpha-z)^2} e^{(\alpha-1)\gamma} = \left(\frac{\gamma e^\tau}{e\tau} \right)^2 (z-1) e^{(z-1)\gamma}. \quad (88)$$

where $\tau = \frac{\phi-z}{2} \gamma \wedge 1$.

Proof of Lemma 15: $D_z(w \| q) - D_1(w \| q)$ is positive because Renyi divergence is an increasing function of the order by [35, Lemma 9-(a)]. In order to bound $D_z(w \| q) - D_1(w \| q)$ from above we use Taylor's theorem. Let $g(\alpha)$ be

$$g(\alpha) \triangleq \int \left(\frac{dw}{dq} \right)^\alpha q(dy).$$

Then $g(\alpha)$ is continuous in α on $(0, \phi)$ by [11, Corollary 2.8.7] because $\int \left(\frac{dw}{dq} \right)^\phi q(dy) = e^{(\phi-1)D_\phi(w \| q)} < \infty$ by the hypothesis and $\left(\frac{dw}{dq} \right)^\alpha \leq 1 + \left(\frac{dw}{dq} \right)^\phi$. Furthermore, using the derivative test we can bound $x^\alpha |\ln x|^\kappa$ as follows

$$x^\alpha |\ln x|^\kappa \leq \left(\frac{\kappa}{e\alpha} \right)^\kappa \mathbb{1}_{\{x \in [0, 1]\}} + \left(\frac{\kappa}{e(\phi-\alpha)} \right)^\kappa x^\phi \mathbb{1}_{\{x \in (1, \infty)\}} \quad \forall \alpha \in (0, \infty), \quad \forall \kappa \in \mathbb{N}.$$

Hence,

$$\left| \frac{d}{d\alpha} \left(\frac{dw}{dq} \right)^\alpha \right| = \left(\frac{dw}{dq} \right)^\alpha \left| \ln \frac{dw}{dq} \right|^\kappa \leq \left(\frac{\kappa}{e\alpha} \right)^\kappa \mathbb{1}_{\left\{ \frac{dw}{dq} \leq 1 \right\}} + \left(\frac{\kappa}{e(\phi-\alpha)} \right)^\kappa \left(\frac{dw}{dq} \right)^\phi \mathbb{1}_{\left\{ \frac{dw}{dq} > 1 \right\}} \quad \forall \alpha \in (0, \infty), \quad \forall \kappa \in \mathbb{N}. \quad (89)$$

The expression on the right hand side has a finite integral for any $\alpha \in (0, \phi)$. Thus as a result of [11, Corollary 2.8.7], κ^{th} derivative of $g(\alpha)$ is given by

$$\frac{d^\kappa}{d\alpha^\kappa} g(\alpha) = \int \left(\frac{dw}{dq} \right)^\alpha \left(\ln \frac{dw}{dq} \right)^\kappa q(dy) \quad \forall \alpha \in (0, \phi), \quad \kappa \in \mathbb{N}. \quad (90)$$

Since $g(\alpha)$ is twice differentiable applying Taylor's theorem [20, Appendix B4] around $\alpha = 1$ we get

$$g(z) \leq 1 + (z-1) \frac{d}{d\alpha} g(\alpha) \Big|_{\alpha=1} + \frac{(z-1)^2}{2!} \sup_{\alpha \in (1, \phi)} \frac{d^2}{d\alpha^2} g(\alpha). \quad (91)$$

Using equation (89) and (90) together with the fact that $g(\alpha) > 1$ for $\alpha \geq 1$ we get³⁴

$$\frac{d^2}{d\alpha^2} g(\alpha) \leq 2 \left(\frac{2}{e(\phi-\alpha)} \right)^2 g(\phi) \quad \forall \alpha \in (1, \phi). \quad (92)$$

Then using the identity $\ln x \leq x - 1$ together with equations (90), (91) and (92) we get

$$\ln g(z) \leq (z-1) D_1(w \| q) + (z-1)^2 \left(\frac{2}{e(\phi-z)} \right)^2 g(\phi) \quad \forall z \in (1, \phi).$$

Then using the identity $g(\alpha) = e^{(\alpha-1)D_\alpha(w \| q)}$ together with the hypothesis $D_\phi(w \| q) \leq \gamma$ we get,

$$D_z(w \| q) - D_1(w \| q) \leq \frac{4(z-1)}{e^2(\phi-z)^2} e^{(\phi-1)D_\phi(w \| q)} \leq \frac{4(z-1)}{e^2(\phi-z)^2} e^{(\phi-1)\gamma}. \quad (93)$$

Note that $D_\alpha(w \| q) \leq \gamma$ for any $\alpha \in (z, \phi)$ because $D_\phi(w \| q) \leq \gamma$ and the Renyi divergence is an increasing function of the order by [35, Lemma 9-(a)]. Thus using the analysis leading to equation (93) we get

$$D_z(w \| q) - D_1(w \| q) \leq \frac{4(z-1)}{e^2(\alpha-z)^2} e^{(\alpha-1)\gamma} \quad \forall \alpha \in (z, \phi]. \quad (94)$$

Using the derivative test we can confirm that the least upper bound among the upper bounds given in (94) is the one at $\alpha = \phi \wedge \left(\frac{2}{\gamma} + z \right)$ and the resulting upper bound is the one given in (88). Note that the least upper bound is less than the upper bound at $\alpha = \phi$ iff $\gamma(\phi - z) > 2$. \blacksquare

B. Renyi Divergence Tradeoff for Tilting and The Dummy Channel Method

Arguably the idea of using the performance of a code on a dummy channel as an anchor to bound its performance is as old as the information theory itself. In a nutshell, dummy channel method can be described as follows: given a channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ and a code (Ψ, Θ) ,

- (i) Choose a function $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ based on the channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ and the code (Ψ, Θ) .
- (ii) Bound the performance of the codes of the form $(f \circ \Psi, \Theta)$.
- (iii) Bound the performance of (Ψ, Θ) using the bound derived in part (ii) and properties of f .

Dummy channel method is employed implicitly in almost all outer bounds for channel coding problems. As an example let us derive Arimoto's outer bound given in [35, Equation (17)]. Given an (M, L) code on a \mathcal{W} such that $C_{1, \mathcal{W}} < \ln \frac{M}{L}$,

³⁴Using equation (90) and bounds similar to (92) one can show that $g(\alpha)$ is not only infinitely differentiable but also analytic on $(0, \phi)$.

let α be the order such that $E_{sp}(\ln \frac{M}{L}, \mathcal{W}) = \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - \ln \frac{M}{L})$. Let f be a constant function such that $f(w) = q_{\alpha, \mathcal{W}}$ for all w in \mathcal{W} . Let P_e^f be the average error probability of $(f \circ \Psi, \Theta)$, then $P_e^f = (1 - \frac{L}{M})$ for all (M, L) codes (Ψ, Θ) . Furthermore, by the monotonicity of the Renyi divergence in the underlying σ -algebra, i.e. [35, Lemma 9-(f)], we have $d_\alpha(P_e^{av} \| P_e^f) \leq D_\alpha(p \circ \mathcal{W} \| p \circ (f(\mathcal{W})))$. On the other hand, $D_\alpha(p \circ \mathcal{W} \| p \otimes q_{\alpha, \mathcal{W}}) \leq C_{\alpha, \mathcal{W}}$ by [35, Theorem 1]. Thus $d_\alpha(P_e^{av} \| P_e^f) \leq C_{\alpha, \mathcal{W}}$. Then [35, Equation (17)] follows from $\frac{\alpha \ln(1-P_e^{av})}{\alpha-1} - \ln(1-P_e^f) \leq d_\alpha(P_e^{av} \| P_e^f)$.

Haroutunian [29] applied the dummy channel method to bound error probability of codes on DSPCs with feedback from below. The exponential decay rate of Haroutunian's bound with block length, however, is greater than the sphere packing exponent for certain channels. Lemma 16 in the following establish the existence of dummy channels with certain desirable features for all channels that has compact closure in the topology of setwise convergence. In the following section, we use these dummy channels to obtain a lower bound on the error probability whose exponential decay rate with block length is equal to the sphere packing exponent.

Lemma 16. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a channel such that $\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}} \neq C_{1, \mathcal{W}}$ and $\mathcal{W} \prec^{uni} \nu$ for some $\nu \in \mathcal{P}(\mathcal{Y}, \mathcal{Y})$.*

(a) *$\forall R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$ there exists a $\phi \in (0, 1)$ and a $z \in (\phi, 1)$ such that*

$$R = C_{\phi, \mathcal{W}} \quad \text{and} \quad E_{sp}(R, \mathcal{W}) = \frac{1-z}{z} C_{z, \mathcal{W}}. \quad (95)$$

(b) *$\forall R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$, $\forall w \in \mathcal{W}$ there exists an $\alpha \in [\phi, z]$ such that³⁵*

$$D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| q_{\alpha, \mathcal{W}}) \leq R \quad \text{and} \quad D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| w) \leq E_{sp}(R, \mathcal{W}) \quad (96)$$

where $v_\alpha^{w, q_{\alpha, \mathcal{W}}}$ is the order α tilted probability measures for w and $q_{\alpha, \mathcal{W}}$ defined in equation (28).

(c) *$\forall R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$, $\forall \epsilon \in (0, \phi)$, $\forall w \in \mathcal{W}$ there exists an $\alpha \in [\phi, z]$ such that³⁶*

$$D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| q_{\alpha, \mathcal{W}}^\epsilon) \leq R + \frac{\epsilon}{(\phi-\epsilon)(1-z)} R \quad \text{and} \quad D_1(v_\alpha^{w, q_{\alpha, \mathcal{W}}} \| w) \leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{\phi(\phi-\epsilon)(1-z)} R \quad (97)$$

where $v_\alpha^{w, q_{\alpha, \mathcal{W}}^\epsilon}$ is the order α tilted probability measures for w and $q_{\alpha, \mathcal{W}}^\epsilon$ defined in equation (28).

(d) *$\forall R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$, $\forall \epsilon \in (0, \phi)$, there exists a function $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ such that³⁷*

$$C_{x, f(\mathcal{W})} \leq R + \frac{\epsilon R}{(\phi-\epsilon)(1-z)} + \left(\frac{R e^{\tau_x - 1}}{(1-z)(\phi-\epsilon)\tau_x} \right)^2 (x-1) e^{(x-1) \frac{R}{(1-z)(\phi-\epsilon)}} + \ln \frac{\epsilon + (1-\epsilon)(z-\phi)}{\epsilon} \quad \forall x \in (1, \frac{1}{z}) \quad (98)$$

$$D_1(f(w) \| w) \leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{\phi(\phi-\epsilon)(1-z)} R \quad \forall w \in \mathcal{W} \quad (99)$$

where $\tau_x = \frac{R(\frac{1}{z}-x)}{2(1-z)(\phi-\epsilon)} \wedge 1$.

(e) *If $q_{\alpha, \mathcal{W}} = q_{\frac{1}{2}, \mathcal{W}}$ for all $\alpha \in (0, 1)$ then $\forall R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$, there exists a function $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ such that³⁸*

$$C_{x, f(\mathcal{W})} \leq R + \left(\frac{R e^{\tau_x - 1}}{(1-z)\phi\tau_x} \right)^2 (x-1) e^{(x-1) \frac{R}{(1-z)\phi}} \quad \forall x \in (1, \frac{1}{z}) \quad (100)$$

$$D_1(f(w) \| w) \leq E_{sp}(R, \mathcal{W}) \quad \forall w \in \mathcal{W} \quad (101)$$

where $\tau_x = \frac{R(\frac{1}{z}-x)}{2(1-z)\phi} \wedge 1$.

Proof of Lemma 16:

(16-a) $C_{\alpha, \mathcal{W}}$ is continuous in α on $(0, 1]$ by [35, Lemma 11-(c)]. Then for any $R \in (\lim_{\alpha \downarrow 0} C_{\alpha, \mathcal{W}}, C_{1, \mathcal{W}})$ there exists a $\phi \in (0, 1)$ such that $R = C_{\phi, \mathcal{W}}$ by the intermediate value theorem [42, 4.23].

As a result of Lemma 3

$$E_{sp}(R, \mathcal{W}) = \sup_{\alpha \in [\phi, 1]} \frac{1-\alpha}{\alpha} (C_{\alpha, \mathcal{W}} - R) > 0. \quad (102)$$

Furthermore $E_{sp}(R, \mathcal{W}) < \frac{1-\phi}{\phi} C_{\phi, \mathcal{W}}$ because $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is decreasing in α by [35, Lemma 11-(c)]. On the other hand, $\lim_{\alpha \uparrow 1} \frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}} = 0$ by [35, Lemma 21-(d)] because $\mathcal{W} \prec^{uni} \nu$ by the hypothesis. Then the existence of z follows from the intermediate value theorem, [42, 4.23], because $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is continuous by [35, Lemma 11-(c)].

(16-b) $q_{\alpha, \mathcal{W}}$ is continuous in α by [35, Lemma 17]. Thus we can replace $q_{\alpha, \mathcal{W}}^\epsilon$ with $q_{\alpha, \mathcal{W}}$ in the proof of part (c) to prove equation (96).

(16-c) We denote $v_\alpha^{w, q_{\alpha, \mathcal{W}}^\epsilon}$ by v_α in order to avoid notational clutter. Since $C_{\alpha, \mathcal{W}}$ is increasing in α by [35, Lemma 11-(a)] and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is decreasing in α on $(0, 1)$ by [35, Lemma 11-(c)], we can bound $\tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ defined in equation (43) as follows:

$$\tilde{C}_{\alpha, \mathcal{W}}^\epsilon \leq C_{\alpha, \mathcal{W}} + \frac{\epsilon}{(1-\epsilon)} \frac{C_{\alpha, \mathcal{W}}}{\alpha(1-\alpha)} \quad \forall \alpha \in (0, 1). \quad (103)$$

³⁵The constant α depends on \mathcal{W} , R and w ; we suppress those dependencies in our notation.

³⁶The constant α depends on \mathcal{W} , R , ϵ and w ; we suppress those dependencies in our notation.

³⁷Both the function f and the constant τ_x depend on \mathcal{W} , R and ϵ ; we suppress those dependencies in our notation.

³⁸Both the function f and the constant τ_x depend on \mathcal{W} and R ; we suppress those dependencies in our notation.

Since $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ is decreasing in α by Lemma 11-(c), using equation (102) we get,

$$E_{sp}(R, \mathcal{W}) \leq \frac{1-\phi}{\phi} C_{\phi, \mathcal{W}} = \frac{1-\phi}{\phi} R. \quad (104)$$

$D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon$ by Lemma 10 and $D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) = D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon) + \frac{\alpha}{1-\alpha} D_1(v_\alpha \| w)$ for all $\alpha \in (0, 1)$ by the definition of tilted probability measure given in equation (28). Thus

$$D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon) + \frac{\alpha}{1-\alpha} D_1(v_\alpha \| w) \leq \tilde{C}_{\alpha, \mathcal{W}}^\epsilon \quad \forall \alpha \in (0, 1), w \in \mathcal{W}. \quad (105)$$

Then using the non-negativity of the Renyi divergence for probability measures, i.e. [35, Lemma 9-(g)], we can bound $D_1(v_\phi \| q_{\phi, \mathcal{W}}^\epsilon)$ and $D_1(v_z \| w)$:

$$D_1(v_\phi \| q_{\phi, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\phi, \mathcal{W}}^\epsilon \quad \forall w \in \mathcal{W} \quad (106)$$

$$D_1(v_z \| w) \leq \frac{1-z}{z} \tilde{C}_{z, \mathcal{W}}^\epsilon \quad \forall w \in \mathcal{W}. \quad (107)$$

As a result of equation (106), $D_1(v_\phi \| q_{\phi, \mathcal{W}}^\epsilon)$ and $D_1(v_z \| q_{z, \mathcal{W}}^\epsilon)$ satisfy one of the following three cases:

- (i) If $D_1(v_\phi \| q_{\phi, \mathcal{W}}^\epsilon) = \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ then $D_1(v_\phi \| w) = 0$ by equation (105) and equation (97) holds for $\alpha = \phi$ by equation (103).
- (ii) If $D_1(v_z \| q_{z, \mathcal{W}}^\epsilon) \leq \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ then equation (97) holds for $\alpha = z$ by equations (103), (104) and (107).
- (iii) If $D_1(v_\phi \| q_{\phi, \mathcal{W}}^\epsilon) < \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ and $D_1(v_z \| q_{z, \mathcal{W}}^\epsilon) > \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ then $\exists \alpha \in (\phi, z)$ such that $D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon) = \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ by the intermediate value theorem [42, 4.23] provided that $D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon)$ is continuous in α . In order to establish the continuity of $D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon)$ in α , first note that $\|q_{\alpha, \mathcal{W}}^\epsilon - q_{\tilde{\alpha}, \mathcal{W}}^\epsilon\| \leq \frac{1-\epsilon}{\epsilon} |\alpha - \tilde{\alpha}|$. Thus $q_{\alpha, \mathcal{W}}^\epsilon$ is continuous in α for the total variation topology. Then $D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon)$ is continuous in α by Lemma 7-(b). Furthermore, for the α satisfying $D_1(v_\alpha \| q_{\alpha, \mathcal{W}}^\epsilon) = \tilde{C}_{\phi, \mathcal{W}}^\epsilon$ we have $D_1(v_\alpha \| w) \leq \tilde{E}_{sp}^\epsilon(\tilde{C}_{\phi, \mathcal{W}}^\epsilon, \mathcal{W})$ by equation (105) and the definition of the averaged sphere packing exponent given in (47). Since $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ a decreasing³⁹ function of R using Lemma 11 we get

$$D_1(v_\alpha \| w) \leq \tilde{E}_{sp}^\epsilon(R, \mathcal{W}) \leq E_{sp}(R, \mathcal{W}) + \frac{\epsilon}{(\phi - \epsilon)\phi} R.$$

- (16-d) Let $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ be $f(w) = v_\alpha^{w, q_{\alpha, \mathcal{W}}^\epsilon}$ for all $w \in \mathcal{W}$ where $v_\alpha^{w, q_{\alpha, \mathcal{W}}^\epsilon}$ is the probability measure described in part (c) satisfying (97). Then the inequality given in (99) is satisfied. We show in the following that the inequality given in (98) is satisfied for this choice of f , as well.

The Renyi divergence is an increasing function of the order by [35, Lemma 11-(a)] and $\alpha \in [\phi, z]$ by part (c) then

$$D_{\frac{1}{z}}(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon) \leq D_{\frac{1}{\alpha}}(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon). \quad (108)$$

Using the alternative expression for Renyi divergence given in [35, Equation (25)] we get

$$\begin{aligned} D_{\frac{1}{\alpha}}(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon) &= \frac{1}{\frac{1}{\alpha} - 1} \ln \int \left(e^{(1-\alpha)D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon)} \left(\frac{dw}{d\nu} \right)^\alpha \left(\frac{dq_{\alpha, \mathcal{W}}^\epsilon}{d\nu} \right)^{1-\alpha} \right)^{\frac{1}{\alpha}} \left(\frac{dq_{\alpha, \mathcal{W}}^\epsilon}{d\nu} \right)^{1-\frac{1}{\alpha}} \nu(dy) \\ &\leq D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon). \end{aligned} \quad (109)$$

Using Lemma 10, together with the monotonicity of $C_{\alpha, \mathcal{W}}$ and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}}$ in α , i.e. [35, Lemma 11-(a,c)], we get

$$D_\alpha(w \| q_{\alpha, \mathcal{W}}^\epsilon) \leq \frac{C_{1/2, \mathcal{W}}}{(1-\alpha)(1-\epsilon)} \leq \frac{R}{(1-z)(\phi - \epsilon)}. \quad (110)$$

We can bound $D_{\frac{1}{z}}(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon)$ using equations (108), (109), (110). Then applying Lemma 15 we get,

$$D_x(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon) \leq D_1(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon) + \left(\frac{R e^{\tau_x - 1}}{(1-z)(\phi - \epsilon)^{\tau_x}} \right)^2 (x-1) e^{(x-1) \frac{R}{(1-z)(\phi - \epsilon)}} \quad \forall x \in (1, \frac{1}{z}), \forall w \in \mathcal{W} \quad (111)$$

where $\tau_x = \frac{R(\frac{1}{z} - x)}{2(1-z)(\phi - \epsilon)} \wedge 1$.

Let $\tilde{\epsilon}$ and $\tilde{\alpha}$ be $\tilde{\epsilon} = \epsilon + (z - \phi)(1 - \epsilon)$ and $\tilde{\alpha} = \frac{1-\epsilon}{1-\tilde{\epsilon}}\phi$. Then

$$q_{\alpha, \mathcal{W}}^\epsilon \leq_{\tilde{\epsilon}} q_{\tilde{\alpha}, \mathcal{W}}^{\tilde{\epsilon}} \quad \forall \alpha \in [\phi, z].$$

Then as a result of [35, Lemma 9-(b)] we have

$$D_x(f(w) \| q_{\tilde{\alpha}, \mathcal{W}}^{\tilde{\epsilon}}) \leq D_x(f(w) \| q_{\alpha, \mathcal{W}}^\epsilon) + \ln \frac{\tilde{\epsilon}}{\epsilon} \quad \forall w \in \mathcal{W}. \quad (112)$$

³⁹ $\tilde{E}_{sp}^\epsilon(R, \mathcal{W})$ is decreasing in R because it is the pointwise supremum of functions each of which is decreasing in R .

Equation (98) follows from equations (97), (111), (112) and [35, Lemma 13].

(16-e) Let $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ be $f(w) = v_\alpha^{w, q_\alpha, \mathcal{W}}$ for all $w \in \mathcal{W}$ where $v_\alpha^{w, q_\alpha, \mathcal{W}}$ is the probability measure described in part (b) satisfying (96). Following an analysis analogous to what we have done for part (d), we can verify that this choice of f satisfies both (100) and (101). ■

C. An Outer Bound for Codes on Product Channels With Feedback

We bound the error probability of codes on DSPCs with feedback using the dummy channel method. We choose the dummy channel using Lemma 16-(d), bound the error probability on the dummy channel using Arimoto's outer bound [35, Lemma 5], and bound the error probability in terms of the error probability on the dummy channel using the monotonicity of the Renyi divergence in the underlying σ -algebra [35, Lemma 9-(f)].

While choosing the dummy channel, one is initially inclined to apply Lemma 16-(d) to \mathcal{W} or to the component channels, i.e. \mathcal{W}_i 's for $i \in \mathcal{T}$. Both choices fail to give good bounds because of the approximation error terms that emerge. Instead we apply Lemma 16-(d) to $\mathcal{W}_{\overrightarrow{(t, \tau)}}$'s where $\tau - t \approx \frac{2}{\kappa}$ for an appropriately chosen κ . In [9], while proving a statement similar to Theorem 3, Augustin used subblocks in a similar way; other ingredients of Augustin's analysis, however, are quite different. Palaiyanur discussed Augustin's proof sketch in more detail in his thesis [39, A.8].

Lemma 17. *Let $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ be a DSPC with feedback for the index set $\mathcal{T} = \{1, \dots, n\}$, ϕ be an order in $(0, 1)$, z be an order in $(\phi, 1)$ satisfying⁴⁰ $\frac{1-z}{z} C_{z, \mathcal{W}} = E_{sp}(C_{\phi, \mathcal{W}}, \mathcal{W})$ and M, L, κ be positive integers satisfying*

$$\ln \frac{M}{L} \geq C_{\phi, \mathcal{W}} + \frac{2C_{\phi, \mathcal{W}}}{\phi(1-z)} \left(\epsilon + \frac{\epsilon+2}{\sqrt[3]{\kappa}} \right) + \kappa^{\frac{1}{3}} + \kappa \ln \frac{1}{\epsilon} \quad \text{and} \quad \frac{1}{n} \lfloor \frac{n}{\kappa} \rfloor \geq \frac{4z\phi}{C_{\phi, \mathcal{W}}} \quad (113)$$

for an $\epsilon \in (0, \frac{\phi}{2})$. Then any (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ satisfies

$$\ln P_e^{av} \geq -\frac{1}{1-e^{-\frac{2}{\sqrt[3]{\kappa}}}} \left(E_{sp}(C_{\phi, \mathcal{W}}, \mathcal{W}) + \ln 2 + \frac{\epsilon 2 C_{\phi, \mathcal{W}}}{\phi^2(1-z)} \right). \quad (114)$$

Proof of Lemma 17: We divide the index set \mathcal{T} into κ subblocks; each subblock is either of length $\lfloor n/\kappa \rfloor$ or of length $\lceil n/\kappa \rceil$. In particular, we set t_0 to zero and define ℓ_i and t_i as follows

$$\ell_i \triangleq \lceil n/\kappa \rceil \mathbb{1}_{\{i \leq n - \lfloor n/\kappa \rfloor \kappa\}} + \lfloor n/\kappa \rfloor \mathbb{1}_{\{i > n - \lfloor n/\kappa \rfloor \kappa\}} \quad t_i \triangleq t_{i-1} + \ell_i \quad \forall i \in \{1, \dots, \kappa\}.$$

For the sake of notational brevity we denote $\mathcal{Y}_{(t_{i-1}, t_i]}$ by $\tilde{\mathcal{Y}}_i$ and $\mathcal{Y}_{(t_0, t_i]}$ by $\tilde{\mathcal{Y}}_{(0, i]}$. Evidently $\mathcal{Y}_{(0, n]} = \tilde{\mathcal{Y}}_{(0, \kappa]}$ and $\mathcal{Z}^{\mathcal{Y}}_{(0, n]} = \mathcal{Z}^{\tilde{\mathcal{Y}}}_{(0, \kappa]}$. Furthermore,⁴¹

$$\mathcal{W}_{\overrightarrow{(0, n]}} = \tilde{\mathcal{W}}_{\overrightarrow{(0, \kappa]}} \quad \text{where} \quad \tilde{\mathcal{W}}_i \triangleq \mathcal{W}_{\overrightarrow{(t_{i-1}, t_i]}} \text{ for all } i \in \{1, \dots, \kappa\}.$$

Thus for any $w \in \mathcal{W}$, i.e. $w \in \mathcal{W}_{\overrightarrow{(0, n]}}$, there exist \tilde{w}_i 's satisfying $\tilde{w}_i(\tilde{\mathcal{Y}}_{(0, i)} | \cdot) \in \tilde{\mathcal{W}}_i$ for all $\tilde{\mathcal{Y}}_{(0, i)} \in \tilde{\mathcal{Y}}_{(0, i)}$ and

$$w(\tilde{\mathcal{Y}}_{(0, \kappa]}) = \prod_{i=1}^{\kappa} \tilde{w}_i(\tilde{\mathcal{Y}}_{(0, i)} | \tilde{\mathcal{Y}}_i) \quad \forall \tilde{\mathcal{Y}}_{(0, \kappa]} \in \tilde{\mathcal{Y}}_{(0, \kappa]}. \quad (115)$$

In a sense, for each $i \in \{1, \dots, \kappa\}$ corresponding \tilde{w}_i is a function from $\tilde{\mathcal{Y}}_{(0, i)}$ to $\tilde{\mathcal{W}}_i$, i.e. $\tilde{w}_i : \tilde{\mathcal{Y}}_{(0, i)} \rightarrow \tilde{\mathcal{W}}_i$.

The stationarity of \mathcal{W} , i.e. the stationarity of $\mathcal{W}_{\overrightarrow{(0, n]}}$, and Lemma 2 implies that

$$C_{\alpha, \tilde{\mathcal{W}}_i} = \frac{\ell_i}{n} C_{\alpha, \mathcal{W}}. \quad \forall \alpha \in (0, \infty), \quad \forall i \in \{1, \dots, \kappa\}.$$

Let us first apply Lemma 16-(d) to $\tilde{\mathcal{W}}_i$ for $R = C_{\phi, \tilde{\mathcal{W}}_i}$, i.e. for $R = \frac{\ell_i}{n} C_{\phi, \mathcal{W}}$. Note that $\tau_x = 1$ for all $x \in (1, \frac{1+z}{2z})$ because $\frac{\ell_i}{n} C_{\phi, \mathcal{W}} \geq 4z\phi$ by the hypothesis. Then for each $\forall \epsilon \in (0, \frac{\phi}{2})$, there exists a map $f_i : \tilde{\mathcal{W}}_i \rightarrow \mathcal{P}(\tilde{\mathcal{Y}}_i, \mathcal{Z}^{\tilde{\mathcal{Y}}_i})$ such that

$$C_{x, f_i(\tilde{\mathcal{W}}_i)} \leq C_{\phi, \tilde{\mathcal{W}}_i} + \frac{\epsilon 2 C_{\phi, \tilde{\mathcal{W}}_i}}{\phi(1-z)} + \left(\frac{2 C_{\phi, \tilde{\mathcal{W}}_i}}{\phi(1-z)} \right)^2 (x-1) e^{(x-1) \frac{2 C_{\phi, \tilde{\mathcal{W}}_i}}{\phi(1-z)}} + \ln \frac{1}{\epsilon} \quad \forall x \in (1, \frac{1+z}{2z}) \quad (116)$$

$$D_1(f_i(w) \| w) \leq E_{sp}(C_{\phi, \tilde{\mathcal{W}}_i}, \tilde{\mathcal{W}}_i) + \frac{\epsilon 2 C_{\phi, \tilde{\mathcal{W}}_i}}{\phi^2(1-z)} \quad \forall w \in \tilde{\mathcal{W}}_i. \quad (117)$$

We define the function $f : \mathcal{W} \rightarrow \mathcal{P}(\mathcal{Y}, \mathcal{Y})$ describing the dummy channel using f_i 's as follows

$$f(w) \triangleq \prod_{i=1}^{\kappa} (f_i \circ \tilde{w}_i) \quad \forall w \in \mathcal{W}. \quad (118)$$

⁴⁰The existence of such a z is established in Lemma 16-(a).

⁴¹Both $\mathcal{Z}^{\tilde{\mathcal{Y}}_i}$ and $\tilde{\mathcal{W}}_i$ are finite sets for all $i \in \{1, \dots, \kappa\}$; thus $((\tilde{\mathcal{Y}}_i, \mathcal{Z}^{\tilde{\mathcal{Y}}_i}), \tilde{\mathcal{W}}_i)$ is a discrete channel for all $i \in \{1, \dots, \kappa\}$. The size of $\tilde{\mathcal{W}}_i$ grows rapidly with the length of the subblock ℓ_i : $\ln |\tilde{\mathcal{W}}_i| = \left(\sum_{j=0}^{\ell_i-1} |\mathcal{Y}_1|^j \right) \ln |\mathcal{W}_1|$. This rapid growth would have made our bounds very weak, if not useless, if the approximation error terms in Lemma 16-(d) were given in terms of $|\mathcal{W}|$ rather than $C_{\phi, \mathcal{W}}$. Note that $C_{\phi, \tilde{\mathcal{W}}_i}$ grows only linearly with ℓ_i .

where for each $i \in \{1, \dots, \kappa\}$, $\tilde{w}_i : \tilde{\mathcal{Y}}_{(0,i)} \rightarrow \tilde{\mathcal{W}}_i$ is the function described in equation (115) and $(f_i \circ \tilde{w}_i) : \tilde{\mathcal{Y}}_{(0,i)} \rightarrow f_i(\tilde{\mathcal{W}}_i)$ is its composition with f_i . Hence, the probability that $f(w)$ assigns to each $y \in \mathcal{Y}$, i.e. to each $\tilde{y}_{(0,\kappa]} \in \tilde{\mathcal{Y}}_{(0,\kappa]}$, is given by

$$f(w)(\tilde{y}_{(0,\kappa]}) = \prod_{i=1}^{\kappa} (f_i \circ \tilde{w}_i)(\tilde{y}_{(0,i)} | \tilde{y}_i) \quad \forall \tilde{y}_{(0,\kappa]} \in \tilde{\mathcal{Y}}_{(0,\kappa]}, \quad \forall w \in \mathcal{W}. \quad (119)$$

where $(f_i \circ \tilde{w}_i)(\tilde{y}_{(0,i)} | \cdot) = f_i(\tilde{w}_i(\tilde{y}_{(0,i)} | \cdot))$ for all $\tilde{y}_{(0,i)} \in \tilde{\mathcal{Y}}_{(0,i)}$. Thus $f(\mathcal{W})$ is a discrete product channel⁴² with feedback for the index set $\{1, \dots, \kappa\}$. Using Lemma 2 we get,

$$C_{\alpha, f(\mathcal{W})} = \sum_{i=1}^{\kappa} C_{\alpha, f_i(\tilde{\mathcal{W}}_i)} \quad \forall \alpha \in (0, \infty). \quad (120)$$

Using equations (116) and (120), together with the identities $\ell_i C_{\alpha, \mathcal{W}} = n C_{\alpha, \tilde{\mathcal{W}}_i}$ and $\ell_i \leq \frac{2n}{\kappa}$ we get

$$C_{x, f(\mathcal{W})} \leq C_{\phi, \mathcal{W}} + \frac{\epsilon 2 C_{\phi, \mathcal{W}}}{\phi(1-z)} + \frac{2}{\kappa} \left(\frac{2 C_{\phi, \mathcal{W}}}{\phi(1-z)} \right)^2 (x-1) e^{(x-1) \frac{4 C_{\phi, \mathcal{W}}}{\kappa \phi(1-z)}} + \kappa \ln \frac{1}{\epsilon} \quad \forall x \in (1, \frac{1+z}{2z}). \quad (121)$$

If $x = 1 + \frac{\phi(1-z)}{4 C_{\phi, \mathcal{W}}} \kappa^{\frac{2}{3}}$ then $x \leq \frac{1+z}{2z}$ because $\frac{1}{n} \lfloor \frac{n}{\kappa} \rfloor \geq \frac{4z\phi}{C_{\phi, \mathcal{W}}}$ and $\kappa \geq 1$ by the hypotheses of the lemma. Then,

$$C_{x, f(\mathcal{W})} \leq C_{\phi, \mathcal{W}} + \frac{2 C_{\phi, \mathcal{W}}}{\phi(1-z)} \left(\epsilon + e \kappa^{-\frac{1}{3}} \right) + \kappa \ln \frac{1}{\epsilon} \quad \text{if } x = 1 + \frac{\phi(1-z)}{4 C_{\phi, \mathcal{W}}} \kappa^{\frac{2}{3}}. \quad (122)$$

In order to bound $D_1(f(w) \| w)$ for $w \in \mathcal{W}$, i.e. for $w \in \mathcal{W}_{(0, n]}$, first note that $n E_{sp}(C_{\phi, \tilde{\mathcal{W}}_i}, \tilde{\mathcal{W}}_i) = \ell_i E_{sp}(C_{\phi, \mathcal{W}}, \mathcal{W})$ because $\ell_i C_{\alpha, \mathcal{W}} = n C_{\alpha, \tilde{\mathcal{W}}_i}$ for all $\alpha \in (0, \infty)$. Then using equations (115), (117) and (119) we get

$$D_1(f(w) \| w) \leq E_{sp}(C_{\phi, \mathcal{W}}, \mathcal{W}) + \frac{\epsilon 2 C_{\phi, \mathcal{W}}}{\phi^2(1-z)} \quad \forall w \in \mathcal{W}. \quad (123)$$

Now we are ready to apply the dummy channel method. Let (Ψ, Θ) be an (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$. Then $(f \circ \Psi, \Theta)$ is an (M, L) channel code on $((\mathcal{Y}, \mathcal{Y}), f(\mathcal{W}))$. We denote the average error probability of $(f \circ \Psi, \Theta)$ by P_e^f . Note that $\ln \frac{M}{L} \geq C_{1, f(\mathcal{W})}$ by (121) and the hypothesis given in (113). Using Arimoto's outer bound [35, Lemma 5] and the definition of sphere packing exponent given in (11) we get

$$\begin{aligned} P_e^f &\geq 1 - e^{-E_{sp}(\ln \frac{M}{L}, f(\mathcal{W}))} \\ &\geq 1 - e^{-\frac{1-x}{x} (C_{x, f(\mathcal{W})} - \ln \frac{M}{L})} \quad \forall x \in (1, \infty) \end{aligned} \quad (124)$$

Then using (124) at $x = 1 + \frac{\phi(1-z)}{4 C_{\phi, \mathcal{W}}} \kappa^{\frac{2}{3}}$ together with (122) and the hypothesis given in (113) we get,

$$P_e^f \geq 1 - e^{-\sqrt[3]{\kappa}}. \quad (125)$$

Let p be the p.m.f. generated by the encoder Ψ on \mathcal{W} when each messages has equal probability mass. Then using the monotonicity of Renyi divergence in the underlying σ -algebra [35, Lemma 9-(f)] we get

$$D_1(p \circ f(\mathcal{W}) \| p \circ \mathcal{W}) \geq d_1(P_e^f \| P_e^{av}). \quad (126)$$

On the other hand,

$$\sup_{w \in \mathcal{W}} D_1(f(w) \| w) \geq D_1(p \circ f(\mathcal{W}) \| p \circ \mathcal{W}) \quad (127)$$

Using the fact that $x \ln x + (1-x) \ln(1-x) \geq -\ln 2$ for all $x \in [0, 1]$ we get

$$d_1(P_e^f \| P_e^{av}) \geq -\ln 2 - P_e^f \ln P_e^{av}. \quad (128)$$

Thus using equations (126) (127) and (128) we get

$$\ln P_e^{av} \geq \frac{1}{P_e^f} (-\ln 2 - \sup_{w \in \mathcal{W}} D_1(f(w) \| w)). \quad (129)$$

Then equation (114) follows from equations (123), (125) and (129). ■

⁴² $f(\mathcal{W})$ is stationary iff ℓ_i is same for all $i \in \{1, \dots, \kappa\}$, i.e. iff $\frac{n}{\kappa} \in \mathbb{N}$.

D. Proof of Theorem 3

Proof of Theorem 3: Lemma 17 is stated in terms of a parameter ϕ in $(0, 1)$ with approximation error terms that depend on ϕ . We apply Lemma 17 only for values of ϕ in $[\alpha_0, \alpha_1]$, but we need approximation error terms to be the same for all values ϕ . We can obtain such a lemma by making certain worst case assumptions on the parameters. In the following, we state a modification of Lemma 17 for $\epsilon = n^{-2}$ and $\kappa = \lfloor n^{\frac{3}{4}} \rfloor$ that has the same approximation error terms for all orders in $[\alpha_0, \alpha_1]$. We state the modified Lemma 17, hence its bound on the error probability, in terms of the component channel \mathcal{W}_0 . In order to obtain bounds in terms of $E_{sp}\left(\frac{\ln M_n/L_n}{n}, \mathcal{W}_0\right)$, we derive an upper bound on $E_{sp}(C_{\phi, \mathcal{W}_0}, \mathcal{W}_0)$.

- Let ϕ be an order in $[\alpha_0, \alpha_1]$, z be an order in $(\phi, 1)$ satisfying $E_{sp}(C_{\alpha_1, \mathcal{W}_0}, \mathcal{W}_0) = \frac{1-z}{z} C_{z, \mathcal{W}_0}$, and M, L be positive integers satisfying

$$\ln \frac{M}{L} \geq n C_{\phi, \mathcal{W}_0} + \frac{2C_{\alpha_1, \mathcal{W}_0}}{\alpha_0(1-z)} \left(n^{-1} + 2\epsilon n^{\frac{3}{4}} \right) + n^{\frac{1}{4}} + 2n^{\frac{3}{4}} \ln n \quad \text{and} \quad n \geq \left(\frac{8}{C_{\alpha_0, \mathcal{W}_0}} \right)^4 \vee \sqrt{\frac{2}{\alpha_0}}.$$

Then any (M, L) channel code on $((\mathcal{Y}_{(0, n]}, \mathcal{Z}_{(0, n]}), \mathcal{W}_{(0, n]})$ satisfies

$$\ln P_e^{av} \geq -n E_{sp}(C_{\phi, \mathcal{W}_0}, \mathcal{W}_0) - \frac{e^{-\frac{4}{\sqrt{n}}} n C_{\alpha_0, \mathcal{W}_0}}{(1-e^{-\frac{4}{\sqrt{n}}}) \alpha_0} - \frac{\ln 2}{1-e^{-\frac{4}{\sqrt{n}}}} - \frac{2C_{\alpha_1, \mathcal{W}_0}}{n(\alpha_1)^2(1-z)(1-e^{-\frac{4}{\sqrt{n}}})}. \quad (130)$$

- Let ϕ be an order in $[\alpha_0, \alpha_1]$ and δ be a real number in $(0, C_{1, \mathcal{W}_0} - C_{\phi, \mathcal{W}_0})$. Then as a result of Lemma 3 we have

$$E_{sp}(C_{\phi, \mathcal{W}_0}, \mathcal{W}_0) \leq E_{sp}(C_{\phi, \mathcal{W}_0} + \delta, \mathcal{W}_0) + \frac{\delta(1-\alpha_0)}{\alpha_0}. \quad (131)$$

For n large enough (M_n, L_n) codes satisfying the hypothesis of Theorem 3 satisfy the hypothesis of the modified Lemma 17 for an $\phi \in [\alpha_0, \alpha_1]$ such that

$$\left| \ln \frac{M_n}{L_n} - \left(n C_{\phi, \mathcal{W}_0} + \frac{2C_{\alpha_1, \mathcal{W}_0}}{\alpha_0(1-z)} \left(n^{-1} + 2\epsilon n^{\frac{3}{4}} \right) + n^{\frac{1}{4}} + 2n^{\frac{3}{4}} \ln n \right) \right| \leq n^{\frac{3}{4}} \ln n$$

Then applying modified Lemma 17 and using equation (131) we can conclude that for n large enough

$$\begin{aligned} \ln P_e^{av} \geq & -n E_{sp}\left(\frac{1}{n} \ln \frac{M_n}{L_n}, \mathcal{W}_0\right) - \frac{(1-\alpha_0)}{\alpha_0} \left(\frac{2C_{\alpha_1, \mathcal{W}_0}}{\alpha_0(1-z)} \left(n^{-1} + 2\epsilon n^{\frac{3}{4}} \right) + n^{\frac{1}{4}} + 3n^{\frac{3}{4}} \ln n \right) \\ & - \frac{e^{-\frac{4}{\sqrt{n}}} n C_{\alpha_0, \mathcal{W}_0}}{(1-e^{-\frac{4}{\sqrt{n}}}) \alpha_0} - \frac{\ln 2}{1-e^{-\frac{4}{\sqrt{n}}}} - \frac{2C_{\alpha_1, \mathcal{W}_0}}{n(\alpha_1)^2(1-z)(1-e^{-\frac{4}{\sqrt{n}}})}. \end{aligned}$$

■

V. DISCUSSION AND FUTURE DIRECTIONS

We have established sphere packing bounds with approximation error terms that are polynomial in the block length for a large class of product channels, including all stationary product channels. Our results hold for a large class of non-stationary channels, which might have infinite channel capacity. We have, also, presented a new proof of the sphere packing bound for codes on DSPCs with feedback. In Appendix B, we derive a sphere packing bound with approximation error terms that are polynomial in the duration, for various memoryless Poisson channels.

The sphere packing bound given in Lemma 12 is applicable to a very broad class of product channels, but its approximation error terms are rather crude. In order to derive bounds with better approximation error terms one can assume $C_{\alpha, \mathcal{W}}$ to be finite for an order α greater than one or \mathcal{W} to be stationarity. It seems, one can improve approximation error terms in Lemma 12 without making such assumptions by making a more definitive use of the continuity of the Renyi centers. While deriving Lemma 12 we used the continuity through the convexity of the Renyi divergence in its second argument and the averaging scheme described in equation (41). Our analysis was, however, indifferent towards the rate of change of the Renyi center as a function of the order. If one can establish bounds on the rate of change of the Renyi center with the order then one can modify the proof of Lemma 12, so as to improve the approximation error terms. If [35, Conjecture 1] is correct, it can be used to bound the rate of change of Renyi center with the order.

We have excluded cost constraints from our discussion. Augustin considered memoryless channels with cost constraints in [9, Chapter VII]. It seems deriving sphere packing bounds for the cost constrained cases is a low hanging fruit. Extending the sphere packing bound results to the models with memory seems to demand more effort.

APPENDIX

A. Minimum Sigma-algebra for Information Transmission

While introducing the channel coding problem formally in Definition 2, we have required the decoding function to be a measurable function, but allowed the encoding function to be any function. We have omitted the measurability requirement on the encoding function because it is inconsequential: The domain of the encoding function for the channel coding problem is a finite message set and there is a tacit assumption that the σ -algebra on the message set is the power set of the message set. Hence, every function from the message set is a measurable function.

For other information transmission problems, such as the joint source-channel coding problem, the encoding function might be from an infinite set equipped with a σ -algebra that is weaker than the power set of this infinite set. In that case, there are functions that are not measurable and the set of all feasible encoding functions is curtailed by the measurability requirement. Measurability of a function depends not only on the σ -algebra on its domain but also on the σ -algebra on its range. Thus one needs to specify a σ -algebra on \mathcal{W} in order to talk about measurability of an encoding function. Unless \mathcal{W} is a finite set, however, there is not one but many reasonable choices for the σ -algebra on \mathcal{W} .

The set of all σ -algebras on \mathcal{W} is partially ordered by inclusion. For any two σ -algebras \mathcal{W}_1 and \mathcal{W}_2 on \mathcal{W} , we say that \mathcal{W}_1 is stronger than \mathcal{W}_2 , or equivalently \mathcal{W}_2 is weaker than \mathcal{W}_1 , iff \mathcal{W}_2 is a sub- σ -algebra of \mathcal{W}_1 , i.e. if $\mathcal{W}_2 \subset \mathcal{W}_1$. Evidently, the power set of \mathcal{W} , i.e. $2^{\mathcal{W}}$, is the trivial maximum and $\{\emptyset, \mathcal{W}\}$ is the trivial minimum. More interestingly, it is possible to define a minimum σ -algebra on \mathcal{W} considering the needs of information transmission problems.

In a general information transmission problem, a code is a strategy to convey a message from the transmitter, which is encoding it to a member of \mathcal{W} , to the receiver, which is decoding it from the output events $\mathcal{E} \in \mathcal{Y}$. While choosing a strategy for the information transmission, one would like *to be able to differentiate between the members of the input set \mathcal{W} based on the probabilities they assign to output events $\mathcal{E} \in \mathcal{Y}$* . In other words for every output event $\mathcal{E} \in \mathcal{Y}$ we want the function $f_{\mathcal{E}} : \mathcal{W} \rightarrow [0, 1]$ given by $f_{\mathcal{E}}(w) = w(\mathcal{E})$ to be $(\mathcal{W}, \mathcal{G})$ -measurable where \mathcal{G} is the σ -algebra generated by the sets (intervals) of the form $[0, \gamma)$ for $\gamma \in [0, 1]$, i.e. $\sigma(\{[0, \gamma) : \gamma \in [0, 1]\})$. The weakest σ -algebra \mathcal{W} satisfying this constraint is called the minimum σ -algebra for information transmission.⁴³

Definition 13. For any channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$ the *minimum σ -algebra for information transmission* is a σ -algebra on the input set \mathcal{W} , given by

$$\sigma_{\mathcal{W}} \triangleq \sigma(\{\mathcal{A}_{\mathcal{E}, \gamma} : \forall \gamma \in [0, 1], \forall \mathcal{E} \in \mathcal{Y}\}) \quad \text{where} \quad \mathcal{A}_{\mathcal{E}, \gamma} = \{w \in \mathcal{W} : w(\mathcal{E}) \in [0, \gamma)\} \quad (132)$$

and for any $\mathcal{A} \subset 2^{\mathcal{W}}$, $\sigma(\mathcal{A})$ is the minimum σ -algebra on \mathcal{W} for which \mathcal{A} is a subset.⁴⁴

Note that $\mathcal{B}([0, 1]) = \mathcal{G}$, i.e. $\mathcal{B}([0, 1])$ is generated by the sets (intervals) of the form $[0, \gamma)$ for $\gamma \in [0, 1]$, because of [11, 1.2.11 Lemma]. Thus for any $\mathcal{E} \in \mathcal{Y}$, the inverse image of any Borel subset of $[0, 1]$ for $f_{\mathcal{E}}$ can be written in terms of countable intersections, unions and complements of members of $\{\mathcal{A}_{\mathcal{E}, \gamma} : \gamma \in [0, 1]\}$. Hence, $f_{\mathcal{E}}$ is $(\mathcal{W}, \mathcal{B}([0, 1]))$ -measurable for any σ -algebra \mathcal{W} including $\{\mathcal{A}_{\mathcal{E}, \gamma} : \gamma \in [0, 1]\}$ for any $\mathcal{E} \in \mathcal{Y}$. Consequently $f_{\mathcal{E}}$ is $(\sigma_{\mathcal{W}}, \mathcal{B}([0, 1]))$ -measurable for any $\mathcal{E} \in \mathcal{Y}$.

We have constructed $\sigma_{\mathcal{W}}$ by a constraint that is motivated by the operational needs of the information transmission problems. $\sigma_{\mathcal{W}}$ is the minimum σ -algebra satisfying that constraint, i.e. $\sigma_{\mathcal{W}}$ is a sub- σ -algebra of any σ -algebra satisfying that constraint. Lemma 18, presented in the following, shows that above mentioned constraint is equivalent to a sufficient condition for a σ -algebra to turn a channel into a transition probability. Let us recall the concept of transition probability first.

Definition 14. Let $(\mathcal{Z}, \mathcal{Z})$ and $(\mathcal{Y}, \mathcal{Y})$ be measurable spaces. Then a function $W : \mathcal{Z} \times \mathcal{Y} \rightarrow [0, 1]$ is called a transition probability (stochastic kernel, Markov kernel) from $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$ if it satisfies following two constraints:

- (i) For all $z \in \mathcal{Z}$, the function $W(z|\cdot) : \mathcal{Y} \rightarrow [0, 1]$ is a probability measure on $(\mathcal{Y}, \mathcal{Y})$.
- (ii) For all $\mathcal{E} \in \mathcal{Y}$, the function $W(\cdot|\mathcal{E}) : \mathcal{Z} \rightarrow [0, 1]$ is a $(\mathcal{Z}, \mathcal{B}([0, 1]))$ -measurable function.

We denote the set of all transition probabilities from $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$ by $\mathcal{P}((\mathcal{Z}, \mathcal{Z})|(\mathcal{Y}, \mathcal{Y}))$.

Lemma 18. For any channel $((\mathcal{Y}, \mathcal{Y}), \mathcal{W})$, let $W : \mathcal{W} \times \mathcal{Y} \rightarrow [0, 1]$ be $W(w|\mathcal{E}) = w(\mathcal{E})$. Then W is a transition probability from $(\mathcal{W}, \mathcal{W})$ to $(\mathcal{Y}, \mathcal{Y})$ iff $\sigma_{\mathcal{W}} \subset \mathcal{W}$.

Proof of Lemma 18: For every $w \in \mathcal{W}$, $W(w|\cdot)$ is a probability measure on \mathcal{Y} by the definition of W . Furthermore, $W(\cdot|\mathcal{E}) : \mathcal{W} \rightarrow [0, 1]$ is $(\sigma_{\mathcal{W}}, \mathcal{B}([0, 1]))$ -measurable by [11, 1.2.11 Lemma] and definitions of $\sigma_{\mathcal{W}}$ and W . Thus, if $\sigma_{\mathcal{W}} \subset \mathcal{W}$ then W is a transition probability.

On the other hand, if W is a transition probability then for all $\mathcal{E} \in \mathcal{Y}$ inverse images of the sets of the form $[0, \gamma)$ for $W(\cdot|\mathcal{E})$ are members of \mathcal{W} . Consequently $\mathcal{A}_{\mathcal{E}, \gamma} \in \mathcal{W}$ for all $\mathcal{E} \in \mathcal{Y}$ and $\gamma \in [0, 1]$. Thus $\sigma_{\mathcal{W}}$ —the minimum σ -algebra generated by $\mathcal{A}_{\mathcal{E}, \gamma}$ — is a subset of \mathcal{W} because the minimum σ -algebra generated by a family of sets is the intersection of all σ -algebras including all members of the family. ■

While analyzing the performance of a code the input set \mathcal{W} is relevant only through its influence on the probabilistic behavior of the channel output. Thus while building a probability space to analyze the performance of a code, instead of working with a $(\mathcal{Z}, \mathcal{W})$ -measurable encoding function $\Psi : \mathcal{Z} \rightarrow \mathcal{W}$ and a transition probability W from $(\mathcal{W}, \mathcal{W})$ to $(\mathcal{Y}, \mathcal{Y})$, one might want to work with a transition probability V from $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$. Lemma 19 proves that these two approaches are equivalent provided that the σ -algebra on \mathcal{W} is $\sigma_{\mathcal{W}}$, by showing that the measurability of the encoding function Ψ is a necessary and sufficient condition for V to be a transition probability.

⁴³Definition 13 employs the concept of minimum σ -algebra generated by a family of sets for describing $\sigma_{\mathcal{W}}$. An equivalent, but more direct, approach is using the concept of minimum σ -algebra generated by a family of functions to describe $\sigma_{\mathcal{W}}$: $\sigma_{\mathcal{W}} = \sigma(\mathcal{F})$ for $\mathcal{F} = \{f_{\mathcal{E}} : \mathcal{E} \in \mathcal{Y}\}$. A discussion of the concept of minimum σ -algebra generated by a family of functions can be found in [11, Section 2.12-i].

⁴⁴A proof of the existence and the uniqueness of the minimum σ -algebra containing a family of sets can be found in [11, Proposition 1.2.6].

Lemma 19. Let $(\mathcal{Z}, \mathcal{Z})$ and $(\mathcal{Y}, \mathcal{Y})$ be measurable spaces, \mathcal{W} be a set of probability measures on $(\mathcal{Y}, \mathcal{Y})$ and $V : \mathcal{Z} \times \mathcal{Y} \rightarrow [0, 1]$ be a function such that $\forall z \in \mathcal{Z}, \exists w \in \mathcal{W}$ satisfying $V(z|\cdot) = w(\cdot)$. Then V is a transition probability from $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$ iff there exists a $(\mathcal{Z}, \sigma_{\mathcal{W}})$ -measurable a function $\Psi : \mathcal{Z} \rightarrow \mathcal{W}$ such that $V(z|\cdot) = \Psi(z)(\cdot)$.

Lemma 19 have an alternative interpretation for the definition of channel as a mapping from the input set to a set of probability measures on the output set. The measurable space $(\mathcal{Z}, \mathcal{Z})$ in Lemma 19 can be interpreted as an input space for the channel V and the function Ψ can be interpreted as the internal encoding function of the channel V mapping the channel inputs to the probability measures on the output space. In this framework Lemma 19 is a necessary and sufficient condition for the channel V to be a transition probability. In this context, Lemma 18 can be interpreted as a special case of Lemma 19 where Ψ is the identity map and $(\mathcal{Z}, \mathcal{Z})$ is $(\mathcal{W}, \mathcal{W})$.

Proof: Let us start with proving the if part. Note that $\{w \in \mathcal{W} : w(\mathcal{E}) \in \mathcal{B}\} \in \sigma_{\mathcal{W}}$ for any $\mathcal{E} \in \mathcal{Y}$ and $\mathcal{B} \in \mathcal{B}([0, 1])$ by [11, 1.2.11 Lemma] and the definition of $\sigma_{\mathcal{W}}$. If there exists a $(\mathcal{Z}, \sigma_{\mathcal{W}})$ -measurable function Ψ such that $V(z|\cdot) = \Psi(z)(\cdot)$ for all $z \in \mathcal{Z}$ then $\Psi^{-1}(\mathcal{S}) \in \mathcal{Z}$ for all $\mathcal{S} \in \sigma_{\mathcal{W}}$. Consequently $V(\cdot|\mathcal{E}) : \mathcal{Z} \rightarrow [0, 1]$ is $(\mathcal{Z}, \mathcal{B}([0, 1]))$ -measurable function and V is a transition probability from⁴⁵ $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$.

Now we proceed with proving the only if part. Assume that V is a transition probability from $(\mathcal{Z}, \mathcal{Z})$ to $(\mathcal{Y}, \mathcal{Y})$. Let $\Psi : \mathcal{Z} \rightarrow \mathcal{W}$ be such that $\Psi(z)$ be $V(z|\cdot)$ for each $z \in \mathcal{Z}$. Since V is a transition probability, $V(\cdot|\mathcal{E})$ is $(\mathcal{Z}, \mathcal{B}([0, 1]))$ -measurable functions for each $\mathcal{E} \in \mathcal{Y}$. Then $\Psi^{-1}(\mathcal{A}_{\gamma, \mathcal{E}}) \in \mathcal{Z}$ for each $\gamma \in [0, 1]$ and $\mathcal{E} \in \mathcal{Y}$. Thus for any $\mathcal{S} \in \sigma_{\mathcal{W}}$ we have $\Psi^{-1}(\mathcal{S}) \in \mathcal{Z}$ because $\sigma_{\mathcal{W}}$ is the minimum σ -algebra of $\mathcal{A}_{\gamma, \mathcal{E}}$'s. Consequently Ψ is a $(\mathcal{Z}, \sigma_{\mathcal{W}})$ -measurable function. ■

If $\mathcal{Z} \neq \mathcal{W}$ then some encoding functions that are measurable for $\sigma_{\mathcal{W}}$ are not measurable for the stronger σ -algebras on \mathcal{W} . Hence, one can curtail the set of feasible encoding schemes by strengthening the σ -algebra on \mathcal{W} . Hence, one needs to specify the σ -algebra used on \mathcal{W} while proving outer bounds. It is, however, not clear at this point that it is always possible or desirable to establish the outer bound for the most general case, i.e. for $\sigma_{\mathcal{W}}$ — the weakest reasonable σ -algebra on \mathcal{W} . The physical constraints on the encoder can justify employing a σ -algebra stronger than $\sigma_{\mathcal{W}}$.

B. The Sphere Packing Bound for Codes on Certain Poisson Channels

The Poisson channels with bounded intensity functions $\Lambda^{[T, a, b]}$ described in [35, Example 10] are product channels. Hence, the analysis presented in Section III in general and Theorem 2 in particular apply to them. More interestingly, with only minor modifications, we can use the same analysis to derive the sphere packing bound for memoryless Poisson channels $\Lambda^{[T, a, b, \varrho]}$, $\Lambda^{[T, a, b, \leq \varrho]}$, and $\Lambda^{[T, a, b, \geq \varrho]}$ described in [35, Examples 8 and 9].

Theorem 4. Let $a, b, \varrho, \alpha_0, T$ be nonnegative real numbers satisfying $\varrho \in [a, b]$, $\alpha_0 \in (0, 1)$, and $T \geq \frac{46}{b-a}(1 \vee \frac{1}{45(1-\alpha_0)})$. Furthermore, let \mathcal{W}_T be a Poisson channel of the form $\Lambda^{[T, a, b, \varrho]}$, $\Lambda^{[T, a, b, \leq \varrho]}$, $\Lambda^{[T, a, b, \geq \varrho]}$, or $\Lambda^{[T, a, b]}$ and M, L be positive integers satisfying

$$C_{1, \mathcal{W}_T} \geq \ln \frac{M}{L} \geq C_{\alpha_0, \mathcal{W}_T} + \frac{1.8}{\alpha_0(1-\alpha_0)} + \frac{11.4}{(1-\alpha_0)} \ln[(b-a)T]. \quad (133)$$

Then any (M, L) channel code on \mathcal{W}_T satisfies

$$P_e^{av} \geq \left[16e^2 e^{\frac{T(b-a)}{T(b-a)\alpha_0-1}} ((b-a)T)^{24.3} \right]^{-\frac{1}{\alpha_0}} e^{-E_{sp}(\ln \frac{M}{L}, \mathcal{W}_T)}. \quad (134)$$

In order to prove Theorem 4, we first establish a parametric outer bound similar to the one in Lemma 12. We use the proof of Lemma 12 to establish the parametric bound given in Lemma 20. It is possible to improve the approximation error terms in Lemma 20 by applying Berry-Essen theorem more carefully. Such a modification will improve the prefactor⁴⁶ in Theorem 4.

Lemma 20. Let $a, b, \varrho, \alpha_0, T$ be nonnegative real numbers satisfying $\varrho \in [a, b]$ and $\alpha_0 \in (0, 1)$. Furthermore, let \mathcal{W} be a Poisson channel of the form $\Lambda^{[T, a, b, \varrho]}$, $\Lambda^{[T, a, b, \leq \varrho]}$, $\Lambda^{[T, a, b, \geq \varrho]}$, or $\Lambda^{[T, a, b]}$ and n, M, L be positive integers satisfying $\frac{M}{L} > 16\sqrt{n}e^{\tilde{C}_{\alpha_0, \mathcal{W}}^{\epsilon} + \frac{\gamma_{\kappa}}{1-\alpha_0}}$ for a $\kappa \geq 3$ and an $\epsilon \in (0, 1)$ satisfying $\frac{(n-1)(1-\alpha_0)(1-\epsilon)}{\epsilon} \geq 1$. Then any (M, L) channel code on \mathcal{W} satisfies

$$P_e^{av} \geq \left(\frac{\epsilon e^{-2\gamma_{\kappa}}}{16e^2(1-\alpha_0)n^{3/2}} \right)^{\frac{1}{\alpha_0}} e^{-\tilde{E}_{sp}^{\epsilon}(R, \mathcal{W})} \quad R = \ln \frac{M}{L} \quad (135)$$

where $\tilde{E}_{sp}^{\epsilon}(R, \mathcal{W})$ is defined in equation (47) and γ_{κ} is given by

$$\gamma_{\kappa} \triangleq 3\sqrt[3]{3} \left(\sum_{t=1}^n \left(\frac{(b-a)T}{n} \vee \kappa \right)^{\kappa} \right)^{\frac{1}{\kappa}}. \quad (136)$$

⁴⁵Note that $V(z|\cdot) : \mathcal{Y} \rightarrow [0, 1]$ is a probability measure on $(\mathcal{Y}, \mathcal{Y})$ because $\mathcal{W} \subset \mathcal{P}(\mathcal{Y}, \mathcal{Y})$.

⁴⁶A less important improvement is obtained by applying Lemma 20 for the largest n satisfying $n \ln n \leq (b-a)T$ rather than $n = \lfloor (b-a)T \rfloor$. In the resulting theorem all $T(b-a)$ terms of theorem 4 are replaced by $\frac{T(b-a)}{\ln T(b-a)}$.

The inequality (135) of Lemma 20 can be replaced by the following alternative inequality.

$$P_e^{av} \geq \left(\frac{\epsilon}{16(1-\alpha_0)n^{3/2}} \right) e^{-2\gamma_\kappa} e^{-\tilde{E}_{sp}^\epsilon(R, \mathcal{W})} \quad R = \ln \frac{M}{L} - 2\gamma_\kappa - \ln \frac{16e^2(1-\alpha_0)n^{3/2}}{\epsilon}. \quad (137)$$

Proof of Lemma 20 and equation (137): Let us first consider $\mathcal{W} = \Lambda^{[T, a, b, \varrho]}$ case. We divide the output space $(\mathcal{X}_T, \mathcal{B}(\mathcal{X}_T))$ into n subblocks⁴⁷ of the form $(\mathcal{X}_{(\frac{t-1}{n}T, \frac{t}{n}T]}, \mathcal{B}(\mathcal{X}_{(\frac{t-1}{n}T, \frac{t}{n}T]}))$ for $t \in \mathcal{T}$ where $\mathcal{T} = \{1, \dots, n\}$. For each $t \in \mathcal{T}$, let $q_{\alpha, t}$ be the Poisson process on the time interval $(\frac{t-1}{n}T, \frac{t}{n}T]$ with the intensity $\beta_{\alpha, a, b, \varrho} = \left(\frac{\varrho-a}{b-a} b^\alpha + \frac{b-\varrho}{b-a} a^\alpha \right)^{\frac{1}{\alpha}}$. Note that $q_{\alpha, t}$ is the order α Renyi center for $\Lambda^{[\frac{T}{n}, a, b, \varrho]}$ determined in [35, Example 8], except for the shift in the time axis. Let $q_{\alpha, t}^\epsilon$ be the corresponding averaged Renyi center defined in equation (41), i.e.

$$q_{\alpha, t}^\epsilon = \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} q_{z, t} \lambda(dz).$$

We repeat the analysis we have done for proving Lemma 12 and equation (58) with the following modifications:

- Instead of $q_{\alpha, \mathcal{W}_t}^\epsilon$, we use $q_{\alpha, t}^\epsilon$.
- Instead of Lemma 10 and equation (44), we use the following derivation to establish equation (61)

$$\begin{aligned} D_\alpha(\Psi(m) \| q_{\alpha}^\epsilon) &= \sum_{t \in \mathcal{T}} D_\alpha(\Psi_t(m) \| q_{\alpha, t}^\epsilon) && \stackrel{(a)}{\leq} \sum_{t \in \mathcal{T}} \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} D_\alpha(\Psi_t(m) \| q_{z, t}) \lambda(dz) \\ &= \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} \sum_{t \in \mathcal{T}} D_\alpha(\Psi_t(m) \| q_{z, t}) \lambda(dz) && \stackrel{(b)}{=} \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} D_\alpha(\Psi(m) \| q_{z, \mathcal{W}}) \lambda(dz) \\ &\stackrel{(c)}{\leq} \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} (1 \vee \frac{\alpha}{1-\alpha} \frac{1-z}{z}) D_z(\Psi(m) \| q_{z, \mathcal{W}}) \lambda(dz) && \stackrel{(d)}{\leq} \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} (1 \vee \frac{\alpha}{1-\alpha} \frac{1-z}{z}) C_{z, \mathcal{W}} \lambda(dz) \\ &\stackrel{(e)}{\leq} \tilde{C}_{\alpha, \mathcal{W}}^\epsilon \end{aligned}$$

The inequality (a) follows from the Jensen's inequality and the convexity of the Renyi divergence in its second argument [35, Lemma 9-(d)]. The equality (b) follows from the product structure of $\Psi(m)$ assumed and the product structure of $q_{z, \mathcal{W}}$ established in [35, Example 8]. The inequality (c) follows from the monotonicity of the Renyi divergence established in [35, Lemma 9-(a)] and the identity $\frac{1-\phi}{\phi} D_\phi(w \| q) = D_{1-\phi}(q \| w)$. The inequality (d) follows from [35, Theorem 1]. The equality (e) follows from the definition of the averaged Renyi capacity given in equation (43).

- Instead of equation (65), which follows from Lemma 10, we use the following identity to bound $D_\alpha(\Psi_t(m) \| q_{\alpha, t}^\epsilon)$.

$$D_\alpha(\Psi_t(m) \| q_{\alpha, t}^\epsilon) \stackrel{(a)}{\leq} \frac{1}{\epsilon} \int_{\alpha-\epsilon\alpha}^{\alpha+\epsilon(1-\alpha)} D_\alpha(\Psi_t(m) \| q_{z, t}) \lambda(dz) \stackrel{(b)}{\leq} \frac{b-a}{1-\alpha} \frac{T}{n}.$$

The inequality (a) follows from the Jensen's inequality and the convexity of the Renyi divergence in its second argument [35, Lemma 9-(d)]. The inequality (b) follows from [35, Equation (99)], the bounds on the intensities of $\Psi_t(m)$ assumed and the expression of for the intensity of $q_{z, t}$ given above.

For proving the lemma for $\Lambda^{[T, a, b, \leq \varrho]}$, we follow the same steps for the intensity $\beta_{\alpha, a, b, \varrho \wedge \varrho_{\alpha, a, b}}$ rather than the intensity $\beta_{\alpha, a, b, \varrho}$, where $\varrho_{\alpha, a, b}$ is defined in [35, Equation (110)]. In [35, Example 9] we have determined the Renyi capacity and the Renyi center of $\Lambda^{[T, a, b, < \varrho]}$. They are equal to the corresponding quantities for $\Lambda^{[T, a, b, \leq \varrho]}$ by [35, Lemma 21-(b)]. For proving the lemma for $\Lambda^{[T, a, b, \geq \varrho]}$, we follow the same steps for $\beta_{\alpha, a, b, \varrho \vee \varrho_{\alpha, a, b}}$. The proof of lemma for $\Lambda^{[T, a, b]}$ follows from the fact that $\Lambda^{[T, a, b]} = \Lambda^{[T, a, b, \leq b]} = \Lambda^{[T, a, b, \geq a]}$. ■

Proof of Theorem 4: We prove Theorem 4, using Lemmas 11 and 20. In order to have uniform approximation error terms, we first bound $C_{1, \Lambda^{[T, a, b]}}$ using the expression given in [35, equation (119)].

$$\begin{aligned} C_{1, \Lambda^{[T, a, b]}} &= \left(\frac{\varrho_{1, a, b} - a}{b-a} b \ln \frac{b}{\varrho_{1, a, b}} + \frac{b - \varrho_{1, a, b}}{b-a} a \ln \frac{a}{\varrho_{1, a, b}} \right) T && \stackrel{(*)}{\leq} \left(\frac{\varrho_{1, a, b} - a}{b-a} b \left(\frac{b}{\varrho_{1, a, b}} - 1 \right) + \frac{b - \varrho_{1, a, b}}{b-a} a \left(\frac{a}{\varrho_{1, a, b}} - 1 \right) \right) T \\ &= \left(\frac{(\varrho_{1, a, b} - a)(b - \varrho_{1, a, b})}{\varrho_{1, a, b}} \right) T && \leq (\sqrt{b} - \sqrt{a})^2 T \\ &\leq (b - a) T. \end{aligned} \quad (138)$$

The inequality (*) follows from $\ln x \leq x - 1$ and the fact that $\varrho_{1, a, b} \in [a, b]$.

While applying Lemmas 11 and 20, we can choose the values of n , κ and ϵ using T provided that the hypotheses of the lemmas are satisfied. Let n , κ and ϵ be

$$n = \lfloor (b-a)T \rfloor \quad \kappa = \ln n \quad \epsilon = \frac{1}{(b-a)T}. \quad (139)$$

⁴⁷We can do that because Borel σ -algebra of the product space is equal to the product of the Borel σ -algebras for second countable topological spaces by [20, Theorem 4.1.7]. See [35, Section V-C] for a more detailed description of Poisson processes.

Then by equation (136) we have

$$\gamma_\kappa \leq 4e \ln n \quad \forall T \geq \frac{46}{b-a}. \quad (140)$$

Since $C_{\alpha, \mathcal{W}_T}$ is increasing in α by [35, Lemma 11-(a)] and $\frac{1-\alpha}{\alpha} C_{\alpha, \mathcal{W}_T}$ is decreasing in α on $(0, 1)$ by [35, Lemma 11-(c)], we can bound $\tilde{C}_{\alpha, \mathcal{W}_T}^\epsilon$ using the definition of averaged Renyi capacity given in (43):

$$\tilde{C}_{\alpha, \mathcal{W}_T}^\epsilon \leq \left(1 + \frac{\epsilon}{1-\epsilon} \frac{\alpha^2 + (1-\alpha)^2}{\alpha(1-\alpha)}\right) C_{\alpha, \mathcal{W}_T}.$$

Note that $C_{\alpha_0, \mathcal{W}_T} \leq C_{\alpha_0, A^{[T, a, b]}}$ by hypothesis and $C_{\alpha_0, A^{[T, a, b]}} \leq C_{1, A^{[T, a, b]}}$ by [35, Lemma 11-(a)]. Then using equations (138), (139) and (140) we get

$$16\sqrt{n}e^{\tilde{C}_{\alpha_0, \mathcal{W}_T}^\epsilon + \frac{\gamma_\kappa}{1-\alpha_0}} \leq 16e^{\frac{46}{45\alpha_0(1-\alpha_0)}} n^{\frac{1-\alpha_0+8\epsilon}{2(1-\alpha_0)}} e^{C_{\alpha_0, \mathcal{W}_T}} \quad \forall T \geq \frac{46}{b-a}. \quad (141)$$

The hypothesis Lemma 20 is satisfied if $T \geq \frac{46}{b-a} \vee \frac{46}{(b-a)45(1-\alpha_0)}$. Thus using equation (140) we get

$$P_{\mathbf{e}}^{av} \geq \left(\frac{\epsilon n^{-8\epsilon}}{16e^{2(1-\alpha_0)} n^{3/2}}\right)^{\frac{1}{\alpha_0}} e^{-\tilde{E}_{sp}^\epsilon(\ln \frac{M}{L}, \mathcal{W}_T)} \quad \forall T \geq \frac{46}{b-a} \vee \frac{46}{(b-a)45(1-\alpha_0)}. \quad (142)$$

By Lemma 11, the hypothesis given in equation (133), the monotonicity $C_{\alpha, \mathcal{W}}$ in α [35, Lemma 11-(a)] we have

$$\tilde{E}_{sp}^\epsilon(\ln \frac{M}{L}, \mathcal{W}_T) \leq E_{sp}(\ln \frac{M}{L}, \mathcal{W}_T) + \frac{C_{1, \mathcal{W}_T}}{(\alpha_0 - \epsilon)\alpha_0} \epsilon.$$

Note that $C_{1, \mathcal{W}_T} \leq C_{1, A^{[T, a, b]}}$ by hypothesis. Then using equations (138) and (139) we get,

$$\tilde{E}_{sp}^\epsilon(\ln \frac{M}{L}, \mathcal{W}_T) \leq E_{sp}(\ln \frac{M}{L}, \mathcal{W}_T) + \frac{1}{(\alpha_0 - \epsilon)\alpha_0}. \quad (143)$$

Equation (134) follows from (142) and (143). ■

ACKNOWLEDGMENT

Author would like to thank Fatma Nakiboğlu and Mehmet Nakiboğlu for their hospitality; this work simply would not have been possible without it. Author completed his initial study confirming the soundness of Augustin's method at UC Berkeley in 2012 fall while he was working as a post doctoral researcher under the supervision of Anant Sahai. Author would like to thank Anant Sahai for his mentorship and his endless patience with the ever forthcoming manuscript, Harikrishna R. Palaiyanur for sending him Augustin's manuscript, Imre Csiszár for pointing out Augustin's work [9] at Austin in 2010 ISIT, Güneş Nakiboğlu for his suggestions on the manuscript, and Vincent Tan for pointing out [43, Theorem 35-(b), (469)].

REFERENCES

- [1] Y. Altug and A. B. Wagner. Refinement of the sphere packing bound for symmetric channels. In *Communication, Control, and Computing (Allerton), 2011 49th Annual Allerton Conference on*, pages 30–37, Sept 2011.
- [2] Y. Altug and A. B. Wagner. A refinement of the random coding bound. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 663–670, Oct 2012.
- [3] Y. Altug and A. B. Wagner. Moderate deviations in channel coding. *IEEE Transactions on Information Theory*, 60(8):4417–4426, Aug 2014.
- [4] Y. Altug and A. B. Wagner. Refinement of the random coding bound. *IEEE Transactions on Information Theory*, 60(10):6005–6023, Oct 2014.
- [5] Y. Altug and A. B. Wagner. Refinement of the sphere-packing bound: Asymmetric channels. *IEEE Transactions on Information Theory*, 60(3):1592–1614, March 2014.
- [6] S. Arimoto. An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20, Jan 1972.
- [7] S. Arimoto. On the converse to the coding theorem for discrete memoryless channels (corresp.). *IEEE Transactions on Information Theory*, 19(3):357–359, May 1973.
- [8] U. Augustin. Error estimates for low rate codes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(1):61–88, 1969.
- [9] Udo Augustin. *Noisy Channels*. Habilitation thesis, Universität Erlangen, 852 Erlangen, West Germany, 1978. (<https://goo.gl/8wupWM>).
- [10] Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.
- [11] Vladimir I Bogachev. *Measure theory*, volume 1. Springer-Verlag, 2007.
- [12] M. V. Burnashev and Yu. A. Kutoyants. On the sphere-packing bound, capacity, and similar results for poisson channels. *Problems of Information Transmission*, 35(2):95–111, 1999.
- [13] G. Como and B. Nakiboglu. Sphere-packing bound for block-codes with feedback and finite memory. In *2010 IEEE International Symposium on Information Theory*, pages 251–255, June 2010.
- [14] I. Csiszár. The method of types [information theory]. *IEEE Transactions on Information Theory*, 44(6):2505–2523, Oct 1998.
- [15] Imre Csiszár and János Körner. *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.
- [16] M. Dalai. Lower bounds on the probability of error for classical and classical-quantum channels. *IEEE Transactions on Information Theory*, 59(12):8027–8056, Dec 2013.
- [17] M. Dalai. Classical and classical-quantum sphere packing bounds: Rényi vs kullback and leibler. In *International Zurich Seminar on Communications*, pages 198–202, 2016.
- [18] M. Davis. Capacity and cutoff rate for poisson-type channels. *IEEE Transactions on Information Theory*, 26(6):710–715, Nov 1980.
- [19] R. L. Dobrushin. An asymptotic bound for the probability error of information transmission through a channel without memory using the feedback. *Problemy Kibernetiki*, 8:161–168, 1962.
- [20] Richard M Dudley. *Real analysis and probability*, volume 74. Cambridge University Press, 2002.

- [21] G. Dueck and J. Körner. Reliability function of a discrete memoryless channel at rates above capacity (corresp.). *IEEE Transactions on Information Theory*, 25(1):82–85, Jan 1979.
- [22] C. G. Esseen. On the liapunoff limit of error in the theory of probability. *Arkiv för matematik, astronomi och fysik*, A28(9):1–19, 1942.
- [23] S. L. Fong and V. Y. F. Tan. Strong converse theorems for classes of multimessage multicast networks: A renyi divergence approach. *IEEE Transactions on Information Theory*, 62(9):4953–4967, Sept 2016.
- [24] R. G. Gallager. A simple derivation of the coding theorem and some applications. *IEEE Transactions on Information Theory*, 11(1):3–18, Jan. 1965.
- [25] Robert G. Gallager. *Information theory and reliable communication*. John Wiley & Sons, Inc., 1968.
- [26] A. Guntuboyina, S. Saha, and G. Schiebinger. Sharp inequalities for f -divergences. *IEEE Transactions on Information Theory*, 60(1):104–121, Jan 2014.
- [27] Te Sun Han. *Information-Spectrum Methods in Information Theory*. Stochastic Modelling and Applied Probability (Book 50). Springer Berlin Heidelberg, 2003.
- [28] E. A. Haroutunian. Estimates of the error probability exponent for a semicontinuous memoryless channel. *Problems of Information Transmission*, 4(4):37–48, 1968.
- [29] E. A. Haroutunian. Lower bound for error probability in channels with feedback. *Problems of Information Transmission*, 13(2):36–44, 1977.
- [30] Evgueni A. Haroutunian, Mariam E. Haroutunian, and Ashot N. Harutyunyan. Reliability criteria in information theory and in statistical hypothesis testing. *Foundations and Trends in Communications and Information Theory*, 4(2-3):97–263, 2007.
- [31] M. Hayashi. Information spectrum approach to second-order coding rate in channel coding. *IEEE Transactions on Information Theory*, 55(11):4947–4966, Nov 2009.
- [32] J. Honda. Exact asymptotics for the random coding error probability. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 91–95, June 2015.
- [33] Yu. M. Kabanov. The capacity of a channel of the poisson type. *Theory of Probability & Its Applications*, 23(1):143–147, 1978.
- [34] J. R. Munkres. *Topology*. Prentice Hall Inc., 2000.
- [35] B. Nakiboğlu. Augustin’s method - part i: The renyi center. [abs/1608.02424v2](https://arxiv.org/abs/1608.02424v2), 2016.
- [36] J. K. Omura. A lower bounding method for channel and source coding probabilities. *Information and Control*, 27(2):148 – 177, 1975.
- [37] H. Palaiyanur and A. Sahai. An upper bound for the block coding error exponent with delayed feedback. In *2010 IEEE International Symposium on Information Theory*, pages 246–250, June 2010.
- [38] H. Palaiyanur and A. Sahai. On haroutunian’s exponent for parallel channels and an application to fixed-delay codes without feedback. *IEEE Transactions on Information Theory*, 61(3):1298–1308, March 2015.
- [39] Harikrishna R. Palaiyanur. *The impact of causality on information-theoretic source and channel coding problems*. PhD thesis, EECS Department, University of California, Berkeley, May 2011.
- [40] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, May 2010.
- [41] Y. Polyanskiy and S. Verdú. Arimoto channel coding converse and renyi divergence. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1327 –1333, Oct 2010.
- [42] Walter Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.
- [43] I. Sason and S. Verdú. f -divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, Nov 2016.
- [44] J. Scarlett, A. Martinez, and A. G. i Fabregas. A derivation of the asymptotic random-coding prefactor. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 956–961, Oct 2013.
- [45] J. Scarlett, A. Martinez, and A. G. i Fabregas. Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations. *IEEE Transactions on Information Theory*, 60(5):2647–2666, May 2014.
- [46] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, The, 27(3 and 4):379–423 and 623–656, July and October 1948.
- [47] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp. Lower bounds to error probability for coding on discrete memoryless channels. i. *Information and Control*, 10(1):65–103, 1967.
- [48] A. Yu Sheverdyaev. Lower bound for error probability in a discrete memoryless channel with feedback. *Problems of Information Transmission*, 18(4):5–15, 1982.
- [49] I. G. Shevtsova. An improvement of convergence rate estimates in the lyapunov theorem. *Doklady Mathematics*, 82(3):862–864, 2010.
- [50] Volker Strassen. Asymptotische abschätzungen in shannons informationstheorie. In *Trans. Third Prague Conf. Inf. Theory*, pages 689–723, 1962. (<http://www.math.cornell.edu/~pmlut/strassen.pdf>).
- [51] A. Valembois and M. P. C. Fossorier. Sphere-packing bounds revisited for moderate block lengths. *IEEE Transactions on Information Theory*, 50(12):2998–3014, Dec 2004.
- [52] S. Verdú and Te Sun Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4):1147–1157, Jul 1994.
- [53] A. B. Wagner and V. Anantharam. Feedback, queueing, and the reliability of the ideal poisson channel above capacity. In *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, page 447, June 2004.
- [54] G. Wiechman and I. I. Sason. An improved sphere-packing bound for finite-length codes over symmetric memoryless channels. *IEEE Transactions on Information Theory*, 54(5):1962–1990, May 2008.
- [55] A. D. Wyner. Capacity and error exponent for the direct detection photon channel. i. *IEEE Transactions on Information Theory*, 34(6):1449–1461, Nov 1988.
- [56] A. D. Wyner. Capacity and error exponent for the direct detection photon channel. ii. *IEEE Transactions on Information Theory*, 34(6):1462–1471, Nov 1988.